

Open Research Online

The Open University's repository of research publications and other research outputs

Parameter Optimization of Conceptual Hydrological Models

Thesis

How to cite:

Eeles, Charles William Owen (1995). Parameter Optimization of Conceptual Hydrological Models. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1994 Charles William Owen Eeles



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000fb77>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

Charles William Owen EELES

**PARAMETER OPTIMIZATION
OF
CONCEPTUAL HYDROLOGICAL
MODELS**

*Date of submission: 23 June 1994
Date of award: 8 June 1995*

A thesis submitted for the degree of Doctor
of Philosophy of the Open University in the
discipline of Applied Mathematics

Date of submission 22nd. June 1994

ProQuest Number:27701207

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 27701207

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT.

The form of modelling used in this research for the simulation of the rainfall/runoff regime of catchment areas by mathematical models is of particular importance to civil engineers in the building of dams, river bridges and other works affected by high and low flows in rivers and streams. The parametric conceptual models can also be used in the management of water resources and as a basis for the assessment of long term risks associated with water storage and transmission of supplies. The objectives of this research are to examine the problems arising from the conceptual modelling of catchment areas with large data sets, and the effective determination of model parameters using gradient and non-gradient optimization techniques in the field of hydrology.

A simple model package was developed from the application and modification of ideas current at the time which allowed a good fit to observed hydrographs to be achieved with the input of rainfall data and data for an evaporation loss function. Nine parameters were available for optimization in this model. The practical demand for the assessment of land use and its variations on catchment water yield led to the development of a more complex model with thirty five parameters based on the latest vegetation process studies.

One of the first modifications was to the criterion for convergence where it was changed from the rate of change of parameter values to that of the model coefficient of determination or efficiency of fit. The least squares

objective function was investigated, and retained for model explained variance. However, for parameters involved in the simulation of base flows it was found to be more effective to use a proportional function, whilst for intense storm events an eighth power function exaggerated the information available in the data for determination of surface runoff parameters. The models employ an input data 'overlay' technique which allowed the use of large data sets running over many years. The simulation results from land use changes with large data sets from the highlands of Scotland, a clay catchment in Buckinghamshire and montane rain forest in Kenya are compared and contrasted for both models.

The results for these catchments using gradient and non-gradient optimization algorithms are also examined, including the use of a genetic algorithm, and recommendations made for the values of algorithm parameters. Hybridized algorithms are developed and tested. A combination of the Rosenbrock and Nelder and Mead Simplex techniques was found to be an efficient hybrid; particularly with the land use model.

PREFACE.

a. Objectives of the Thesis.

1. To assess and compare the suitability of existing algorithms commonly used for the optimization of parameters in models used to simulate hydrologic processes with large sets of data.
2. To consider the effectiveness of other algorithms not commonly used in the solution of such problems with large data sets.
3. To assess the effectiveness of hybrid algorithms.
4. To compare and contrast the simulations obtained by two parametric conceptual models developed with nine parameters and thirty five parameters respectively in their mathematical representation.
5. To assess the effects on the simulation of catchments with different vegetation cover using these models and the optimization process.

b. The importance of modelling in hydrology.

The relationship of the environmental model to an instrumented catchment is one of process simplification, and this has a crucial bearing on the successful extrapolation of the simulation results to other areas. The instrumented catchment is set up either as a representative basin for the investigation of the hydrologic cycle, with the long term purpose of extrapolating the results in time and space, or for

Thesis: C.W.O.Eeles

the study of the changes brought about by anthropogenic effects.

The models are used in the management of water resources, the forecasting of the effects of changes in land use on stream flow, real time flow forecasting with drought or flood warning, and the quality control and infilling or extension of flow data sets for stochastic analyses. In early work on the models considerable problems were experienced with the optimization of their parameters, and the large data sets still provide a difficult problem for the commonly used optimization algorithms.

c. Acknowledgements.

Part of this work was carried out whilst the author was employed by the Natural Environment Research Council at the Institute of Hydrology with the permission and support of the Director of the Institute, and uses hydrologic data collected and processed by the old Experimental Catchments Section, and the Surface Water Archive Section.

The author had some very interesting and stimulating discussions on modelling with Jim Blackie, Dr. Ann Calver and other colleagues at the Institute during this research.

The first internal supervisor was Dr. John Berry, now Professor at the University of the South West, who was mainly responsible for the commencement of the research in the Department of Applied Mathematics and Computing at the Open University. He was followed by Dr. Mick Bromilow when he left to take up an appointment at Plymouth. The external supervisor was Dr. Graham Raggett of the University of Hallam, Sheffield,

Thesis: C.W.O.Eeles

for the whole of the long research period. The author is greatly indebted to the supervisors for their constant inspiration, patience and friendly support without which this work could not have been completed.

The author also has a great debt owed to his family, Angela and Corinna Eeles, who have lived under the shadow of the 'Thesis' for so long and given their love and support without stinting in some very difficult and depressing times.

CONTENTS.

	Page
1.0 Introduction.	
1.1 General approach to modelling of Hydrological Systems.	1.
1.2 The river basin or catchment area as a system.	5.
1.3 Model design and selection criteria.	8.
1.4 Identifying objectives and use of model simulations.	14.
1.5 Problems of snow melt routing.	15.
1.6 Catchment characteristics and field data.	17.
1.7 Accuracy of simulation.	19.
2.0 Development of parametric conceptual models.	
2.1 Early history of the parametric model.	22.
2.2 The first parametric models.	25.
2.3 Early work on the application of optimization algorithms.	28.
2.4 The St. Lucia model.	33.
2.5 HYdrological Rainfall/RunOff Model (HYRROM).	33.
2.6 The Land Use Model.	36.
3.0 Application of conceptual models to hydrologic problems and streamflow forecasting.	
3.1 Objectives of catchment simulations.	38.
3.2 Quality control and infilling of time series data.	39.
3.3 Record extension and synthetic record generation.	41.

3.4	Water resources assessment.	44.
3.5	Management of water resources.	46.
3.6	Future developments in the use and extension of parametric conceptual models.	47.
4.0	The parametric conceptual models.	
4.1	The generalised model concepts.	50.
4.2	The HYdrological Rainfall/RunOff Model	52.
4.3	Structure and concepts of the Land Use Model.	59.
4.4	The complete Land Use Model.	64.
5.0	Model fitting techniques.	
5.1	Systematic approach to parameter adjustment.	70.
5.2	The split record and objective functions.	73.
5.3	Sensitivity of parameters and response surfaces.	76.
5.4	Strategy of parameter optimization.	81.
5.5	Problems arising from the model structure.	84.
5.6	Feasible function space.	85.
5.7	Constraints, scaling and criteria for convergence.	86.
5.8	Mathematical formulation of the search technique.	88.
6.0	Optimization algorithms available in the field of Hydrology.	
6.1	Availability of algorithms.	90.
6.2	The concept of the direct search algorithm.	91.
6.3	The Gram-Schmidt orthogonalization of axes.	94.
6.4	The Rosenbrock rotation of axes algorithm.	96.

Thesis: C.W.O.Eeles

6.5	Step-wise application of the Rosenbrock algorithm.	98.
6.6	The Nelder and Mead Simplex algorithm.	102.
6.7	The Levenberg and Marquardt algorithm.	107.
6.8	NLFIT Package.	109.
7.0	Algorithms new to the field of hydrology.	
7.1	The Numerical libraries.	112.
7.2	The conjugate direction algorithm.	115.
7.3	The Evolution Strategy algorithm of Rechenberg.	120.
8.0	Data used for algorithm tests and model comparison.	
8.1	Catchment data and preliminary analyses.	129.
8.2	Data from the Kirkton catchment, Balquhiddy, Scotland.	131.
8.3	Data from the Ray catchment, Grendon Underwood, Buckinghamshire.	132.
8.4	Data from the Lagan River, montane rain forest catchment, Kericho, Kenya.	133.
8.5	Double mass analyses of data.	134.
8.6	Frequency distributions of data.	138.
8.7	River stage recession curves.	142.
8.8	Plots of river recession curves and regressions.	144.
9.0	Comparison of algorithms and catchment models.	
9.1	Algorithms and data.	159.
9.2	Multimembered Evolution strategy KORR.	164.
9.3	The Rosenbrock algorithm.	165.
9.4	Direct Search algorithm.	165.

Thesis: C.W.O.Eeles

9.5	Nelder and Mead Simplex algorithm.	167.
9.6	The Conjugate Direction algorithm.	170.
9.7	The Marquardt algorithm.	171.
9.8	Results from the tests.	173.
9.9	Direct ranking of algorithms.	175.
9.10	Hybridization of two algorithms.	176.
9.11	Comparison of the two models.	178.

10.0 Conclusions and recommendations for further work.

10.1	The current relevance of the research.	206.
10.2	The problem of global optimization.	208.
10.3	The limits and successes of HYRROM.	211.
10.4	The land use model.	212.
10.5	The sets of test data.	213.
10.6	General problems with the optimization algorithms.	215.
10.7	Results of the tests.	216.
10.8	Hybridization of algorithms.	218.
10.9	Recommendations based on the tests.	218.
10.10	Recommendations for future work.	219.

REFERENCES.

(a).	Published work and reports by C.W.O.Eeles.	221.
(b).	General references.	223.

APPENDICES.

1.	The Penman index of evapotranspiration.	235.
2.	FORTRAN77 code for the HYRROM model	237.
3.	Subroutine MUTATE.	247.

FIGURES

Figure Number	Page Number
2.1 Simulation of the Kenwyn River at Truro.	35.
3.1 Double mass curve of the rainfall for two experimental catchments at Balquhiddar.	43.
3.2 Percentage reduction in runoff for afforestation of the Balquhiddar catchments.	45.
4.1 Schematic of the conceptual model.	51.
4.2 Schema of Parameter Specification of HYRRROM MODEL.	53.
4.3 Schematic of LAND USE MODEL.	60.
4.4 Schematic of LAND USE SUB-MODULE.	65.
5.1 General schema for split record calibration and validation of a model,	71.
5.2 Three dimensional mapping of F-space for the two routing delay parameters.	80.
6.1 Diagram of the first iteration of the Rosenbrock algorithm for two parameters.	95.
6.2 Rosenbrock alternate parameter and alternate axes search.	99.
6.3 Rosenbrock alternate parameter final iteration F-value v. function calculations.	100.
6.4 Rosenbrock alternate parameter final iteration F-values v. computer processing time.	100.
6.5 Rosenbrock alternate axes search final iteration F-values v. number of function calculations.	101.
6.6 Rosenbrock alternate axes search final iteration F-values v. total CPU time.	101.
6.7 Operation of the Nelder and Mead Simplex algorithm	103.
6.8 Simplex search conditions.	104.

7.1	Conjugate direction strategy.	118.
7.2	Rechenberg (2,4) evolution strategy for one generation.	121.
7.3	Schematic of the Rechenberg evolutionary algorithm subroutines and functions.	126.
8.1	Double mass analyses of the Kirkton catchment data.	135.
8.1(a)	Kirkton catchment data 1983 to 1988.	135.
8.2	Double mass analyses of the Ray catchment data.	136.
8.2(a)	Ray catchment data 1964 to 1980.	136.
8.3	Double mass analyses of the Lagan, Kenya montane forest catchment.	137.
8.3(a)	Lagan catchment data 1958 to 1974.	137.
8.4	Histogram of catchment rainfall data.	138.
8.5	Histogram of catchment flow data.	139.
8.6	Histogram of catchment Penman evaporation data.	140.
8.7	Histogram of seasonal mean rainfall data.	140.
8.8	Histogram of seasonal mean streamflow data.	141.
8.9	Histogram of seasonal mean Penman E_0 data.	142.
8.10	Kirkton catchment twelve day recession curves.	145.
8.11	Ray catchment ten day recession curves.	145.
8.12	Lagan, Kenya twenty two day recession curves.	146.
8.13	Kirkton catchment average twelve day recession curve.	147.
8.14	Ray catchment average 10 day recession curve.	147.
8.15	Lagan, Kenya average twenty two day recession curve.	148.
8.16	Lagan, Kenya forty five day recession curve.	148.
8.17	Lagan, Kenya smoothed forty five day recession	149.

curve.

9.1	Histogram of F-values reached by each algorithm for Kirkton catchment.	166.
9.2	Histogram of F-values reached by each algorithm for Ray catchment.	166.
9.3	Histogram of F-values reached by each algorithm for Lagan, Kenya catchment.	167.
9.4	Histogram of Kirkton simulations correlation with observed data.	167.
9.5	Histogram of Ray simulations correlation with observed data.	168.
9.6	Histogram of Lagan, Kenya simulations correlation with observed data.	168.
9.7	Histogram of Kirkton simulations coefficient of variation with observed data.	169.
9.8	Histogram of Ray simulations coefficient of variation with observed data.	169.
9.9	Histogram of Lagan, Kenya simulations coefficient of variation with observed data.	170.

TABLES

Table number	Page number
1.1 Interaction of the system components.	3.
4.1 Table of parameters for both models.	67.
8.1 Kirkton catchment monthly rainfall, streamflow and Penman E_0 .	150.
8.2 Kirkton catchment daily frequency distributions 1984 to 1988.	151.
8.3 Ray catchment monthly rainfall, streamflow and Penman E_0 .	152.
8.4 Ray catchment daily frequency distributions 1964 to 1980.	154.
8.5 Lagan, Kenya monthly rainfall, streamflow and Penman E_0 .	155.
8.6 Lagan, Kenya daily frequency distributions 1964 to 1973.	157.
9.1 Initial values of parameters as set in HYRRROM modelling package.	163.
9.2 EVOL (scaled) algorithm Kirkton results.	180.
9.3 EVOL (scaled) algorithm Ray results.	181.
9.4 EVOL (scaled) algorithm Lagan, Kenya results.	182.
9.5 ROSENBROCK (scaled) algorithm Kirkton results.	183.
9.6 ROSENBROCK (scaled) algorithm Ray results.	184.
9.7 ROSENBROCK (scaled) algorithm Lagan, Kenya results	185.
9.8 DIRECT SEARCH (scaled) algorithm Kirkton results.	186.
9.9 DIRECT SEARCH (scaled) algorithm Ray results.	187.
9.10 DIRECT SEARCH (scaled) algorithm Lagan, Kenya results.	188.
9.11 SIMPLEX (unscaled) algorithm Kirkton results.	189.

Thesis: C.W.O.Eeles

9.12 SIMPLEX (unscaled) algorithm Ray results.	190.
9.13 SIMPLEX (unscaled) algorithm Lagan, Kenya results.	191.
9.14 CONJUGATE DIRECTION (unscaled) algorithm Kirkton results.	192.
9.15 CONJUGATE DIRECTION (unscaled) algorithm Ray results.	193.
9.16 CONJUGATE DIRECTION (unscaled) algorithm Lagan, Kenya results.	194.
9.17 CONJUGATE DIRECTION (unscaled) algorithm Lagan, results with DFS = 100.	195.
9.18 MARQUARDT (scaled) algorithm Kirkton results.	196.
9.19 MARQUARDT (scaled) algorithm Ray results.	197.
9.20 MARQUARDT (scaled) algorithm Lagan, Kenya results.	198.
9.21 Ranking of algorithms for each data set.	199.
9.22 HYBRID (scaled) restart from Rosenbrock parameters for Kirkton.	200.
9.23 HYBRID (scaled) start from initial parameters.	201.
9.24 HYBRID (scaled) start from initial parameters for Ray.	202.
9.25 HYBRID (scaled) start from initial parameters for Lagan, Kenya.	203.
9.26 Comparison of hybrid algorithm model calibrations.	204.
9.27 Comparison of hybrid algorithm model validations.	205.

GLOSSARY (a) - Function, Variable or constant.

Function, Variable or Constant.	Description.
A	Contributing area.
C_{Mulvaney}	The Mulvaney constant.
E_{actual} or AET	Actual evapotranspiration.
E_0	Penman open water evaporation index.
E_{obs}	Observed data set of potential Penman open water evaporation index.
E_i	Evapotranspiration loss, i , from the model system consisting of evaporation from leaf surfaces, open water, etc., and vegetation transpiration from the soil profile. $1 \leq i \leq 3$.
$E_{\text{intercept}}$	Canopy interception loss.
E_T	Penman evapotranspiration index: E_0 modified to allow for vegetation albedo.
E_{tran}	Calder/Newson vegetation transpiration loss.
F	Objective function.
$F_{i,j}$	Objective function for the i -th iteration and j -th parameter.
F_{init}	Initial value of the objective function.
F_{norm}	Normalized objective function
F_*	Optimal value of the objective function.
f_{opt}	Component of the optimal value of the objective function due to parameter p_i .

Thesis: C.W.O.Eeles

$F(p,R,E)$ The main model function with parameters, p , input data of rainfall, R , and evaporation index, E .

This is partitioned into subsets of functions representing the processes affected by the vegetation at the surface and surface topography, the soil profile below this, and the function set containing the groundwater processes.

F_{relative} Measure of the accuracy or efficiency of fit of the simulated hydrograph.

G_{opt} Sensitivity of parameter p_i at the local or global optimum.

G_{store} Groundwater store contents.

K Transpiration reduction function.

n Number of runoff ordinates.

N_i Degree of fit for the i -th. iteration.

P_i The value of the surface runoff partition function at time t_i .

P_{lower} Lower bound of the parameter p_i .

P_{scaled} Scaled value of the parameter p_i .

P_{upper} Upper bound of the parameter p_i .

Q_{catch} Total simulated streamflow or sum of its elements.

Q_i Streamflow for the interval i .

Q_{intent} Peak flow intensity.

Q_g 'Slow response' groundwater component of Q_{catch} .

Q_0 Observed streamflow.

Q_m Mean observed streamflow.

Q_s 'Rapid response' surface component of Q_{catch} .

Q_{sim} Total simulated flow or its elements.

R Observed rainfall, or precipitation, data.

Thesis: C.W.O.Eeles

R_i	Reduced, or effective, rainfall at level i as the total rainfall, R , is processed through the system with evapotranspiration losses and additions to store contents, or rainfall for the interval i .
R_{intent}	Peak rainfall intensity.
R_{total}	Total catchment rainfall.
S	Store contents.
S_{chan}	Surface channel store contents.
$S_{intercept}$	Contents of vegetation interception store.
S_{smd}	Soil moisture deficit. i.e. Size of soil moisture store is $S_{soil} + S_{smd}$.
S_{soil}	Soil profile store contents.
t_i	Data set time interval i of n daily time intervals, $1 \leq i \leq n$.
V	Model explained variance.
V_0	Variance of observed flow data set Q_0 .
W	Fraction of time interval t when canopy is wet.
β	Calder/Newson transpiration constant.
δ	Calder/Newson interception indice.
ΔG_i	Change in groundwater storage.
ΔS_i	Change in soil moisture storage.
ϵ	Criterion for validation of the model.
γ	Calder/Newson interception constant.
Ω_{conv}	Criterion for convergence.
ϕ	The function angle in radians $\phi(p_i, P_{lower}, P_{upper})$.
θ_r	The angle in radians for the cosine function which limits transpiration with high S_{smd} .

GLOSSARY (b) - parameters of the models.

Parameter name	Description	Remarks
Interception store:-		
p_1	Size of interception store (mm). $0 < p_1 \leq 5$	Detention store for grasslands.
p_2	Penman evaporation factor for interception store. $1 < p_2 \leq 2$	Model designed for Penman data, but will use other evaporation data.
p_3	Interception store contents (mm). $0 \leq p_3 \leq 5$	System initial state.
Surface soil and detention store:-		
p_4	Size of surface detention and soil moisture store (mm). $0 < p_4 \leq 10$	Litter and humic layer.
p_5	Penman transpiration factor. $0.3 < p_5 \leq 1$	Penman E_0 or E_T .
p_6	Store contents (mm). $0 \leq p_6 \leq 10$	System initial state.
Surface runoff store:-		
p_7	Partitioning factor between surface runoff and infiltration. $0 < p_7 \leq 1$	Surface runoff estimated and residual is infiltration.
p_8	Exponential factor relating surface runoff to soil moisture deficit. $0 < p_8 \leq 1$	Reduction factor.
p_9	Exponential factor relating surface runoff to intensity of precipitation input. $0 < p_9 \leq 1$	Increase factor.
Channel routing store:-		
p_{10}	Channel routing store exponential. $1 < p_{10} < 3$	Determines shape of release curve.

Thesis: C.W.O.Eeles

P ₁₁	Channel routing store factor. $0.1 < p_{11} \leq 1$	
P ₁₂	Runoff delay. $p_{12} \geq 0$	In model time intervals.
P ₁₃	Initial volume in transit (mm). $p_{13} > 0$	System initial state.
P ₁₄	Initial contents of channel routing store (mm). $p_{14} > 0$	System initial state.

Soil moisture store:-

P ₁₅	Transpiration ceases at this deficit (mm). $p_{16} < p_{15} < 200$	Wilting point.
P ₁₆	Transpiration starts to decrease at this deficit (mm). $0 \leq p_{16} < 50$	Moisture stress point.
P ₁₇	Percolation to groundwater factor. $0 < p_{17} \leq 1$	Percolation only occurs when soil profile is above 'fieldcapacity'.
P ₁₈	Initial soil moisture deficit (mm). $0 < p_{18} < 50$	System initial state.
P ₁₉	Maximum deficit below forest cover when transpiration ceases (mm). $80 \leq p_{19} < 200$	Optimized value removes store contents bias.
P ₂₀	Maximum deficit below heather when transpiration ceases (mm). $40 < p_{20} < 80$	- as above -

Groundwater store:-

P ₂₁	Denominator of groundwater store contents fraction. $30 < p_{21} < 300$	Equal to store contents, G_s , when changing at the rate of 1 mm per time interval.
P ₂₂	Groundwater store exponential. $1 \leq p_{22} < 4$	Sets curvature of recession curve.
P ₂₃	Groundwater delay. $p_{23} \geq 0$	plus 1 data time period.

**Calder/Newson empirical parameters:-
(Process studies values in brackets).**

P_{24}	Sets canopy saturation limit. $0.01 < p_{24} < 0.07$ (0.045)	Limits value for interception loss.
P_{25}	Forest transpiration factor. $0.1 < p_{25} < 1$ (0.910)	
P_{26}	Forest interception factor. $1 < p_{26} < 10$ (6.990)	
P_{27}	Forest interception exponential factor. $0.001 < p_{27} < 0.2$ (0.099)	
P_{28}	Heather transpiration factor. $0.1 < p_{28} < 1$ (0.500)	
P_{29}	Heather interception factor. $1 < p_{29} < 5$ (2.650)	
P_{30}	Heather interception exponential factor. $0.1 < p_{30} < 0.5$ (0.360)	

Catchment vegetation and open water area proportion factors:-

A_g	Grassland.
A_f	Coniferous Forest.
A_d	Deciduous Forest
A_h	Heather moorland.
A_w	Open water.

where $A_g + A_f + A_d + A_h + A_w = 1$
and $0 \leq A \leq 1$.

GLOSSARY (c) - Hydrological terms.

anthropogenic effects: the effects of man on the environment.

basin or catchment: an area drained by a river and its tributaries.

cumec: cubic metre per second.

cumec days: average cumecs per day.

depth over the catchment: depth of a volume of water over the area of a catchment.

domain: an area in which the aspect, slope and altitude are similar.

evapotranspiration: a combination of the evaporation from a particular area and the transpiration from the vegetation in that area.

field capacity: the amount of water held in a given soil matrix by capillarity forces against drainage by gravity.

flume: a shaped channel in which measurements are taken to estimate the volume of water passing through it in unit time.

offtake: a channel, passage or pipe for abstracting water from a river.

piezometric tube: tube with regularly spaced holes for emplacement in the soil and aquifer to measure the piezometric head which is the sum of the pressure and elevation head shown by the level of water in the tube.

sublimation: the direct conversion of water from its solid state to the vapour phase.

Thiessen polygon: a method of weighting the catch of rain gauges by the area of a polygon about the gauge. This polygon is constructed by bisecting a line between it and the next gauge, and then constructing a polygon from the points where the bisectors meet each other and the catchment boundary.

volume fit: the closeness of the simulated flow volume total to the observed flow total over a period of time.

CHAPTER 1.

Introduction.

1.1 General approach to modelling of Hydrological Systems.

The system in nature ('real world') to be simulated by a model is the basin, or catchment area, of a stream or river, the flows in which form the output of the system as a water resource. The system response to the impulse provided by precipitation in the form of rain and snow is compounded with evaporation and transpiration 'loss' processes from the area. This response is not unique for a particular input, but is dependent on the initial state of the system and on the action of the loss processes which are functions of the meteorological input from incident radiation and of the advected energy.

The rainfall/runoff hydrological models were divided into three classes by Wheater (1993). However, this was a very simplistic classification which had little relevance as a means of identifying the relative usefulness of groups of models. The three classes used to group models ran from the very simple through the conceptual to the very large model which attempts to represent 'reality'. Unfortunately, no hydrological model should be completely 'black box' just as no model completely represents the physics of the environmental prototype. Elements of the three 'classes' are present as a model gains in complexity and any classification used becomes indistinct and blurred, and therefore it becomes difficult to 'type' a model.

As a model increases in the amount of information from the environment contained in the model structure design, so does its data requirements and use of computer resources, and in

principle the range of hydrologic problems to which it can be applied. With the increasing complexity there is not necessarily any greater accuracy or advancement of scientific understanding of the processes involved and their relationships. Models are abstractions from a very complex 'real world' and their usefulness (or otherwise) in achieving an objective is the only measure of their success.

The type of model with which this research is concerned is the conceptual model which relies on the amount of information which can be extracted from its input variables to assess the value of its parameters. The early development of this type of model is described in Chapter 2. The concepts of the model are assumed to have a simple structure and their parameters are physically relevant to large sub-areas of the catchment or to the whole area. This makes the parameters conceptually meaningful, but not in general directly measurable. They have to be identified by optimization, either subjectively by trial and error or by experience, or objectively by seeking the minimum value of an error function which generates an n -dimensional parametric hyperspace.

The set of optimal values for the conceptual model is in general not unique, and as will be shown later in this work different optimization algorithms find widely differing sets of parameters giving locally optimal solutions. Problems with the structure of the response surface lead to poor performance by some algorithms which are dependent on calculating the slope of the n -dimensional space.

The theory of modelling and simulation is discussed by Zeigler (1976) and the interactions of his components 'real system', 'model', and 'computer' are equivalent to the three

columns (one, three, and five) shown in Table 1.1 as 'real world' data input, 'computer processing' and 'computer output'. The various operations are indicated by the arrows in columns two, three and four. The arrows in column two show the input of control data from the first column to the model parameter optimization algorithm in the third column which controls the calibration of the model, and also from the first column the initial parameters customizing the model. The input of rainfall and 'loss' function daily data for each month to the model from column one, together with the input of observed streamflow data to the objective function for the same month.

Table 1.1
Component interaction.

'Real world' data input		Computer processing		Computer output
Input control data	→	Optimization	<->	Simulation
		↓		
Rainfall and 'Loss' function data	→	Model	↑	
		↓		
Observed streamflow data	→	Objective function	↑	

The arrows of the fourth column show the return path to the first row where the double arrow indicates the point of decision to either continue the optimization of model parameters, or use the calibrated model to simulate the observed flows.

The model is a dynamic representation of continuous hydrological processes which act on discrete observed input

data; usually at hourly, daily or monthly intervals. The latter are an aggregation of either hourly or daily data as the timing of input and loss events within the month seriously affects the simulation during periods of rapid change. This timing of events within or over several time intervals also affects the error in the simulated response for the daily frequency data. This daily data is also biased by errors in the observed data introduced by instrumentation, networks and observers. With the complicated system found in nature the simulated response is generally unlikely to match exactly the response of the prototype, and has therefore to be within an acceptable limit for the objective of the project:-

$$| \text{Observed data} - \text{Simulated data} | < \epsilon \quad (1.1)$$

where ϵ is a previously chosen criterion for the precision of the simulation.

It is because of the extreme complexity within such systems that it is necessary to consider as a whole the components of the modelling process as in Table 1.1: the parametric model, the observed data to which it is fitted, and the parameter optimization algorithm which aligns the model concepts with the reality of these data. In the second column two components of the computer processing are used to generate a simulated data set which is then compared with 'real' data in the third component - the objective function. The optimization algorithm forms an indispensable link between the model generated output and the observed data. The number of parameters for the models considered in this thesis range from nine to thirty five: with the smaller number of parameters the model is used to simulate the response of a catchment as a

whole while the latter looks at the land use within the area. The parameters are not necessarily independent and there is generally not a unique optimal solution. The physical reality of the prototype is controlled by fixing upper and lower bounds on the parameters to reduce the size of the 'feasible region'. Even then, different sets can be found which give approximately the same degree of fit. The process of optimization is by no means 'automatic' in the hydrologic case!

The design of the model is dependent on the objective of the simulation and the frequency of the data available. The algorithm, used for automatic optimization, is often the only one available at the time, and little thought is given to its suitability for the particular project. If there is a choice then it becomes a matter of personal preference and experience rather than an operationally based option. The main purpose of this thesis is to examine and test the algorithms commonly available at this time in the field of Hydrology, and make recommendations on their suitability for use with the lumped conceptual type of model. These are highly parameterised models using relatively long periods of data for calibration and validation of the fitted model using the usual 'split' records to fit the simulation and test it with different sequences of input data. A further purpose is to ascertain if other available algorithms would be appropriate for the parametric solution of this type of model.

1.2 The river basin or catchment area as a system.

The catchment, or river basin, has been the basic research area of hydrology since Perrault and Marriotte showed in the seventeenth century that the stream flow in the upper

Seine could be attributed solely to the rainfall within that area (Rodda, 1976). However, the ideas and observations behind the hydrological cycle are even older, Leonardo da Vinci (1508-1519), in the ancient world Herodotus (c. 445 BC) uses experiences from his own travels and gives some of those of Thales of Miletus (c. 548 BC). The combination of the thoughts of practical men with observations of their environment in ancient times is a fascinating study.

The area of the catchment is defined by the topographical elements which contribute to the movement of water through the catchment to the lowest point on the boundary called the outfall of the catchment. Here for hydrologic purposes these flows can be measured by a gauging structure or calibrated river section. This gauging point is therefore the key point of the catchment, and is chosen because of its position and suitability in defining the land area that is of interest to the particular project.

This area of land is one from which it is physically possible for the incoming precipitation to form part of the flow from the catchment. The size of the area may range from a hectare or less to millions of square kilometres such as the Amazon basin. The proportion of the incident precipitation that is not lost to the atmosphere as evaporation and transpiration follows a complex network of flow paths between different 'stores' to the outfall point of the catchment. The time over which the actual transport of the moisture takes place can vary considerably: from minutes to decades depending on the size and physical characteristics of the catchment. The upper limit of the output is set by the volume of the precipitation, but this is reduced to an 'effective' amount of

input by the losses within the catchment through direct evaporation, transpiration from the vegetation, and changes in the volume held in various storages. The incident radiant and advected energy by atmospheric movement determines the potential volume loss from the system, but in reality this varies greatly according to a number of factors affecting the actual loss arising from a given potential; varying combinations of these factors can produce the same loss. This 'loss' function has a complicated physical basis which is explained in detail in Appendix 1 where we describe the Penman index of potential evaporation from open water.

There are several factors which contribute to water movement and loss from a catchment. The most obvious factor is the infiltration rate of the surface soil; rain falling on a deep porous sand will quickly pass through it, whilst impervious clay causes water pounding. In the latter case water can only move into the soil through cracks and root paths; the cracks forming from the extremely high adsorptivity of the clay 'platelets' and their relative movement. A somewhat less obvious factor is the vegetation species which is of great practical importance in estimating losses. There vegetation contributes to the loss by rainfall interception on its leaves and branches leading to direct evaporation, and by drawing water from the soil through its root system which is then transpired through the leaf surfaces at a rate dependent on the energy available and on the species. The transpiration process continues as long as the roots can extract water from the surrounding soil, and it will usually contribute more to the water loss from the system than direct evaporation.

The water loss to the system involves the following factors:-

1. The energy available: radiation and advected energy.
2. The volume and intensity of rainfall; its spatial and temporal distribution.
3. The vegetation species; their density and distribution.
4. The soil types; their distribution, infiltration rates and storage capacities.

The remaining water in the system, or 'effective rainfall', not affected by these factors is potentially available to contribute to the stream flow at the catchment outfall. Whether it does so within minutes, days or decades depends on which of the many flow paths through the catchment it is constrained to follow. At one extreme the bulk of the precipitation during a high intensity storm will contribute to the overland flow because the infiltration rate into the soil is exceeded; this is the 'rapid' response flow output from the system. This volume of water moves directly over the surface to the dendritic system of stream channels leading to the catchment outfall in a very short time. At the other extreme is a catchment with slight gradients and very permeable soils where the surface water infiltrates quickly down through the soil profile and into shatter belts and faults in the rock strata forming groundwater aquifers. This movement is slow through these stores, and the water eventually emerges into the stream to form the 'slow' response or 'base flow'. This is a component of the flow which is always present except in extreme drought conditions or where there is little storage within the catchment.

1.3 Model design and selection criteria.

To design a model as an abstract of the complete hydrological system of a catchment which accurately reflects the physical and biological processes governing water movement needs a very detailed knowledge of the system and the way in which these processes interact. Although no model is able to even begin to simulate these processes with the heterogeneity of the 'real world' there would appear to be merit in using the physically based distributed models because of the greater physical relevance of their equations which are applied to each node of a grid covering the catchment. Unfortunately that is exactly what these models are: point estimations of the physical processes. At best it has to be assumed that the parameters governing the processes are smoothly changing between nodes - at worst that they apply to the complete area around the node, and are discontinuous only at the grid mesh boundaries. The question of sensitivity and the effect of the simplification of physical processes is examined in Gustard et al (1989) in relation to the Hupselse Beek research basin in the Netherlands. This particular study applied one of the well developed distributed models and found considerable problems with its application.

There is also a question of physical scale and grid size; difficulties became apparent when attempts were made to apply surface trend analysis to hydrologic problems. Change the scale of a physical process and/or vary the grid size and totally different surfaces appear which are related only at the points common to the generation of each surface. An interesting discussion of discretization and grid scale

demonstrating some of the difficulties which can arise is given by Calver and Wood (1989). These aspects of a distributed model are such that they can give a fortuitously good set of simulated data or prevent a satisfactory simulation.

There are a number of intercomparisons published by the World Meteorological Organisation between models: conceptual models used in operational hydrological forecasting (1975), models of snowmelt runoff (1986), simulated real-time intercomparison of hydrological models (1987), and hydrological models for water resource system design and operation (1990). These studies give a wide perspective of the types of model and their effectiveness when operating on similar problems.

Because of the distributed nature of conceptual models it is impossible to relate processes at a particular position within the basin occurring at some earlier point in time to the combination of the flows at the outfall, unless there is some areal link which allows the processes within a catchment to be integrated or lumped for a significant area.

Since the distributed model has a 'reality' of its own given by its physical equations or sub-models, eg Darcy's Law (a transport equation) or the Richard's equation for the diffusion of soil moisture, it is often regarded as mirroring physical reality and offering an insight into the processes involved. Unfortunately, this very process of finite distribution actually makes the model a 'lumped' model. A critique examining the problems in making practical hydrological predictions using these models is given by Beven (1989), who pays particular regard to the problems associated with physical parameters determined at a point in a heterogeneous system and their representation over a grid area.

The initial programme of field calibrations and determinations of physical parameters in the laboratory are likely to take a considerable time before the results of such measurements can be mounted in a model at the grid scale, which leads to the consideration of an alternative and less expensive approach in which the time could be spent in obtaining time series data from flows under differing conditions for use in calibrating a much simpler conceptual model. Whilst it should be possible in principle to construct a water information system on a geographic basis for ease of input of data and parameters, it appears unlikely that the spatial or vertical resolution of such a data base would ever be sufficient for input to a realistic distributed model. The data is simply not available at this scale; it would seem a pointless exercise to mount a soils map plotted from a 'ground truth' observation every 10 km, on 50 m grid intervals! Even the use of satellite imagery, or aerial photography, is not likely to provide a solution to the problem of the areal extent to which particular physical parameters apply - a relatively simple task such as the mapping of the distribution of vegetation within a catchment is not easily carried out by these techniques, unless there are clearly defined physical boundaries to each vegetation type. Over a large area of mixed vegetation the dominant spectral signature for a particular pixel varies not only with the albedo but with the angle of incident light. The problems of determining areas of vegetation are shown in the paper by Roberts, Johnson and Law (1992).

The question of grid size and discretization is further examined by Calver and Wood (1989) by comparing the effects of changes on simulated flows from a model simulated hillslope

using physically realistic parameters. The widely different discharges per metre of hillslope over time for simulations using different grids are of particular interest in the context of simulating physical reality! Guide lines are laid down for the cost effective use of computer resources in achieving a reasonable accuracy, and the paper shows the considerable problems that occur in this type of modelling which can prevent achievement of results that are robust in their relation to processes in the field.

These models are very labour intensive in their requirements for data entry at points over an extensive catchment area. A further point which emerged during discussion of the keynote paper by Eeles, Robinson and Ward (1990) at the University of Wageningen was that such models are already bounded in their practical applications by the computer resources required to run them over sufficiently long simulation periods. Studies are being made of the applications of parallel processing to such models as reported in several papers in Farmer and Rycroft (1991).

It is suggested that the way forward from the highly physically based type of model is by the use of semi-distributed models using some of the concepts developed in conceptual models at the catchment scale. These parametric models provide areal estimates of the different processes with a relatively simple mathematical representation of their input and output to stores and thus simulating the path of 'effective' rainfall to the catchment outfall. They have proved remarkably successful in estimating the effect of land use changes on the operational use of water resources, Eeles and Douglas (In preparation), and the effect of clear felling

of woodland over several years on low flows as shown in the FRENCH Report by Gustard et al (1989). These models use the most common simplification of lumping, or spatial averaging, as used in the highly distributed models at the grid mesh scale, but at the catchment or sub-catchment scale. The simplifications necessary for small lumped models are discussed in Blackie and Eeles (1985) when applied to the physical structure and processes operating within a catchment. For spatial averaging at the catchment scale the implication is that the whole system can be represented mathematically using only the dimensions of depth and time. In such a system no account is taken within its boundaries of variations of precipitation, soils, geology or topography. To apply this to an extreme case the model input does not differentiate between a 10 mm input of rain occurring uniformly over the whole catchment, or a 30 mm input from a convective storm over one third of the total area. This spatial averaging must occur in all catchment models, including the most complex distributed models, and the only criterion of its success is the achievement of the objective of the simulation given by equation 1.1.

The concepts of the parametric conceptual models have previously been considered as adequate only for small catchments with homogeneous vegetation, soils and geology, but their application has been made to large complex catchment areas of the order of 2,000 Km² such as the Upper Thames river. The variation of flows due to the changes of land use have been simulated using these models operating for each area of use, and weighting the resultant outputs in time as described in Eeles and Douglas (In preparation). The key factors in the successful application of these models is the quasi-steady

state of the catchment system, and stable spatial distributions of precipitation over the catchment topography; even highly asymmetric patterns of rainfall are acceptable so long as the recurring patterns are reasonably stable. The apparent stationarity of processes within a catchment system has been questioned by Robinson, Eeles and Ward (1990), using these models operating at hourly intervals to demonstrate the variation with time and intensity of actual against simulated flows when changes in soil drainage at the surface layer have been made by 'mole' drainage schemes.

From the above discussion it would appear that conceptual models are simple and flexible in use and represent a simple and pragmatic approach to a simulation of catchment response with minimum cost in staff and computer resources. It was originally thought when the large distributed models were first introduced, Abbott et al (1986), that the length of observed data series requirements for calibration of the lumped models often precluded their use. However, it has since become obvious that the time required for the determination of parameters for a fully distributed model is of the same order, but is expensive in terms of field and laboratory work and involves the extremely tedious entry of the results into a large data base.

1.4 Identifying objectives and use of model simulations.

The lumped conceptual models used in this thesis are limited by the simplicity of concept in that their equations represent the integrated processes operating over a whole

catchment, a sub-catchment or particular area of interest. The model does not consider, in detail, the processes at a point and so cannot be used if detailed soil profile work is required for the study of conditions leading to soil movement or erosion within a catchment. The structure of the model has to be defined in terms of the requirements of the study for which it is to be used. For example, what terms of the water balance, equation 1.2, is the model required to simulate, and what is the time interval of interest in considering the processes?

The water balance equation of continuity on the time interval of the data, i , which is used or implied by continuous volume accounting models is:-

$$Q_i = R_i - (E_{\text{actual}})_i - \Delta S_i - \Delta G_i \quad (1.2)$$

Where Q_i = streamflow;
 R_i = precipitation;
 E_{actual} = actual evapotranspiration;
 ΔS_i = change in soil moisture storage;
 ΔG_i = change in groundwater storage.

All variables have units of millimetres over the catchment area, or volume e.g. cumec days, the average cubic metres per second for a period of 24 hours.

Having identified the objective(s) of the study then the question of available data becomes crucial to the choice of model structure. If hourly outputs are required then hourly or shorter interval data for input to the model must be available - the distribution of daily or monthly interval data at shorter time intervals is fraught with problems in terms of the 'Loss' functions and simulated hydrograph.

1.5 Problems of snow melt routing.

The observed precipitation data over a period of days can be 'corrected' for total volume when it is in the form of snowfall, and this has been demonstrated by Johnson et al (1990) in their paper on the Balquhiddar catchments' data. Their distribution of the precipitation equally over the snow period does cause a 'step' in the input data with a consequent broader and lower peak in the model simulated hydrograph. However, although there are satisfactory models of the accumulation and 'ripening' of the snow pack, Morris (1985), WMO (1986), there does not appear to be a model linked to normal climatic data which will give the temporal distribution of snow melt. A further problem is the routing of snow melt: depending on the frozen condition of the surface soil - what is the proportion of snow melt which forms surface runoff and how much infiltrates into the soil?

The snow data for Balquhiddar consists of a note of days when snow was lying in the catchments. There are no details of the extent, depth and density of snow cover. The only temperature data available are those recorded by the automatic weather stations, and a study needs to be carried out to relate these to the temporal distribution of snow melt and its routing to the catchment outfall.

Snow accumulation and melt causes difficulties in the simulated hydrograph as shown for the Monachyle catchment at Balquhiddar from December 1983 through to the end of April 1984. This period is discussed by Eeles and Blackie (1993). The land use model simulation responds immediately to the input of snow and over-predicts the runoff, and when snow melt occurs there is an under-prediction of runoff. Over the full period of the snow events the simulated volume of runoff is in error

by +11% although the time distribution is radically out. This volume error may have been caused by an underestimate of the evaporation and sublimation from the snow pack or by an overestimate of precipitation. Subsequent re-assessment of the 1983 precipitation estimates (Blackie, 1993) indicated that the Monachyle value for December 1983 may have been overestimated by some 40 mm. This is sufficient to account for the overestimate in flow simulated by the land use model.

1.6 Catchment characteristics and field data.

The hydrologic and physical characteristics of a catchment, touched on in section 1.1, are very different according to geographical location, topography and land use. Some catchments are hydrologically degraded e.g. heavily eroded with the erosion material still in the catchment forming extra storage, and with the base rock exposed in the upper areas. This gives a very rapid response to precipitation which is then 'damped' by the erosion material in the valley.

The catchment can be desert (free of vegetation) or can have differing natural vegetation or can have land devoted to various forms of agriculture with intensive use of water resources. Varying degrees of urbanisation can affect surface storage and speed of stream response to rainfall.

For an experimental basin set up in the natural environment and instrumented to obtain observations for a particular objective the networks of instruments can be relatively dense. The network of raingauges will have been designed to sample from each individual domain area identified with the same altitude, aspect and exposure. Automatic weather stations will capture and store meteorological data for later

collection by an observer or direct transmission by satellite. At a central point there will be a manual weather station with twice daily observations. Soil moisture observations will be taken at longer intervals from representative sites in the catchment. A stage recorder will be recording the hydrograph at the outfall of the catchment either continuously by chart or at intervals by analogue recorder. The data will be processed on a computer for later examination, validation and quality control, providing information for the project, and for modelling (which can then be used to extrapolate the results for the management of water resources elsewhere).

Inevitably there will be observer error, instrument error, processing error, and network design error all compounded over the period of the field work by lack of finance, changes of staff, changes of emphasis, computer operating system and data base problems. In general it would appear that for at least the first year of data capture the data will be atypical due to antecedent climate conditions or to subsequent anthropogenic effects. A good model will be a useful quality control tool in identifying areas which need investigation on the database. Care should be taken that it is not inadequacies in the model concepts which are causing difficulties!

At worst the data may be from a raingauge some kilometres away from the catchment with a manual weather station also outside the boundary. An optimization algorithm is then the only means of calibrating the model against the observed flow data. With the lumped conceptual model the concepts of storage are represented by simple storage tanks with connecting links representing the integrated processes,

and it is necessary to optimize the parameters to allow for different catchment characteristics and possible observed data bias.

1.7 Accuracy of simulation.

Due to the complex and heterogenous nature of the prototype there is little likelihood of an exact fit to the observed hydrograph with a unique set of parameters. In the early work it was thought that there existed 'global' parameters, but there has been no experimental support for this view and not one global parameter has been identified by subsequent research. Instead feasible bounds for each parameter are available from field work, theoretical considerations, and numerous applications of the models to different catchments.

An interesting example of the change of bounds is the Penman factor p_2 which is used to modify the Penman Open Water potential evaporation index which applies to unit area (see Appendix 1). This was thought to have an upper limit of one for vegetation in the original canopy evaporation work based on the canopy plan area of the plant. This concept neglected the 'leaf' area and interception capability of the vegetation which is greater than the plan area. Later modelling work showed that this limit was greater than one for the surface detention and interception store. The empirical limits set on this factor are $1 < p_2 \leq 5$.

The measure of the accuracy, or efficiency of fit $F_{relative}$ of the simulated hydrograph can be derived from, Nash and Sutcliffe (1970), the explained variance or objective function, F , using the model:-

$$F = \frac{1}{n} \sum_{j=1}^n (Q_o(t_j) - Q_s(t_j))^2 \quad (1.3)$$

The summation is taken over n values of time t_j , where Q_s is the model simulated flow and Q_o is the observed flow at time t_j . This is the general objective function for the optimization model (with variations for the optimization of 'rapid' and 'slow' response parameters during the parameter optimization process).

The data variance, V_o is given by:-

$$V_o = \frac{1}{n} \sum_{j=1}^n (Q_o(t_j) - Q_m)^2 \quad (1.4)$$

where

$$Q_m = \frac{1}{n} \sum_{j=1}^n Q_o(t_j)$$

where Q_m is the mean of the observed flows. The analogous coefficient of determination, or efficiency of fit, $F_{relative}$ is then given by:-

$$F_{relative} = \frac{V_o - F}{V_o} \quad (1.5)$$

This last equation can be very unstable for small values of n .

From the combination of components and their interactions shown in Table 1.1 it would appear that the modelling and simulation system cannot be broken up into its elements for separate examination. The 'real' data, the optimization algorithm, and the model with its objective function form a system which can be used to identify the optimal parameters for

the model which, in turn, can be used to generate the simulation.

The link between theory, measurement and calculation methods is emphasized and described in Fleming (1975), and this gives further weight to the concept of the combined study of the complete system: data input, optimization algorithm and conceptual model.

Chapter 2.

Development of parametric conceptual models.

2.1 Early history of the parametric model.

Catchment modelling has its beginnings in the need to quantify and estimate the effects of storm runoff which came from the demands made by the rapid expansion of civil engineering projects in the nineteenth century. In the design of canals, railways and urban sewer systems the common requirement was some method of estimating this storm runoff. With the rapid expansion of towns and industry it became necessary to develop methods of estimating flows for water abstraction from the rivers, and in predicting the water resources available when assessing the feasibility of new reservoirs.

These models could only be developed as the hydrological cycle became an object of research, and this further required the measurement of rainfall, streamflows and other variables. The variables are usually measured in volumes for engineers, and depths over the catchment for hydrologists. The former need to know the actual yield from a catchment, and the latter the relative yield for comparison with other catchments. The rational formula proposed by Mulvaney (1851) was one of the first 'event' models relating storm runoff peaks to antecedent rainfall. Its simple form is

$$Q_{\text{peak}} = C_{\text{Mulvaney}} \cdot R_{\text{intent}} \cdot A \quad (2.1)$$

which relates the peak storm runoff Q_{peak} to a constant determined from available data, $C_{Mulvaney}$, the rainfall intensity, R_{intent} , and the total contributing area, A . This type of relationship is common to similar empirical expressions developed subsequently and still in use today.

This expression, 2.1, can be shown to be a special case of the Unit Hydrograph model outlined by Sherman (1932), and subsequently developed to the instantaneous unit hydrograph which formed the basis of storm response models by Nash (1957) and Dooge (1959). The current methods of estimating catchment response to individual storm events have evolved from these beginnings, gaining in complexity with each development. The main aim of these models is to produce estimates of flood peaks for engineering purposes, and they do not normally attempt to estimate volume flows continuously.

Models to give a continuous simulation of the hydrograph were of a similar form to the above attempting to express a rainfall/runoff relationship:-

$$Q_i = C_{historic} \cdot R^{D_{historic}} \quad (2.2)$$

Where Q_i is the flow for the interval of time, i , R is precipitation for that interval of time, and the factor $C_{historic}$ is assumed constant together with the power $D_{historic}$, and both are determined from historic flow and rainfall records.

A more complicated form of continuous modelling became possible only after Penman (1948) in the UK and Thornthwaite (1948) in the USA had developed methods of estimating evapotranspiration from agrometeorological data. This allowed the following continuity equation to form the basis of water

balance models which can operate over far greater time periods than the event type of input/output model:-

$$Q_i = R_i - E_{actual} - \Delta S_i - \Delta G_i \quad (2.3)$$

where Q_i and R_i are stream flow and precipitation respectively for the time interval t_i , E_{actual} is actual evapotranspiration for the interval, ΔS_i is the change in soil moisture storage, and ΔG_i is the change in groundwater storage.

This formulation must be present either explicitly or implicitly in any model operating over a period. The water balance model operates independently on the data for each of the time intervals and is not a continuous function. In the form (2.3) it requires as inputs the change in soil moisture and groundwater which are rarely available for periods of less than a month, if at all. For catchments with well defined seasonal rainfall distributions these terms can be neglected between times of similar moisture deficit and stream flow, or where the cumulative difference between precipitation and evapotranspiration is so large that changes in storage are second order effects. This concept led to the use of data divided up into 'water years' to enable water balance analysis to be used without considering changes in storage.

The straightforward continuity equation needs considerable modification in order to predict the fine structure of the streamflow hydrograph at daily or hourly intervals. This must take the form of water movement through the 'stores' under gravitational potential, and routing functions to simulate the flow in channels.

2.2 The first parametric models.

Theoretical developments had progressed by the early 1950's to allow models of the combined processes to be designed. These included developments on the theory of infiltration into the soil (Horton (1930) and Philip (1954)), and on the movement of liquids through porous mediums, Richards (1931), and the application of his earlier work to the storage and movement of water in soils, Richards et al (1956). The routing of flow in open channels and the Saint Venant (1871) equations for flow over a plane surface was investigated in the modelling context, McCarthy (1938), Morris (1983). Work in the late 1950's coincided with the development of computer hardware which could handle the massive numerical work involved in the use of these models.

The most well known of the parametric conceptual models is the Stanford Watershed Model, Linsley and Crawford (1960), which made an enormous quantum leap in the model development field. This model produced estimates of daily flow from daily rainfall inputs using infiltration, unit hydrograph and baseflow recession functions, Crawford and Linsley (1962). Later the model was extended to include changes in soil moisture storage, evapotranspiration estimation and flow routing techniques for hourly flow estimates, Crawford and Linsley (1964,1966). Continuous development of this model resulted in its licensing by Hydrocomp International on a commercial basis as the 'Hydrocomp Simulation Network' containing the model package 'Hydrologic Simulation Program' later described by Johansen et al (1980). However, this model was very highly parameterised, and, although it later included the automatic optimization of four parameters, it is dependent

on the skill and experience of its operator for successful results. The package is still the major conceptual model covering the whole of the hydrologic cycle.

The 'Tank' model developed by Sugawara (1961) was another of the early models which simulated the movement of water through the system by using simple linear reservoirs in series and in parallel. This model gave good results when evapotranspiration was virtually constant, but had problems when used for example in Kenya, where these losses become extremely variable. When reservoirs became empty during a model run it was found that succeeding reservoirs emptied quickly and were not able to refill until the preceding tanks had filled again preventing the model simulating the observed hydrograph.

Some models had their parameters fitted to observed data by iterative trial and error methods; a method still followed in some simulation applications. The success of the method depended on observation and experience to obtain the best fit to the observed data. A model proposed by Dawdy and O'Donnell (1965) used four stores and conventional infiltration and routing functions: the significant advance with this model was that a modification of an automatic optimization routine given in Rosenbrock (1960) was used to obtain optimal values of the parameters. This model appears to be the first with nine parameters in the literature; from subsequent experience this appears to be the minimum number necessary to obtain a reasonable fit to a catchment hydrograph with a correlation greater than 75%.

This line of research into the parametric conceptual model and its applications was extended by the Institute of

Hydrology from 1969 onwards. It was initially based on a model proposed by Nash and Sutcliffe (1970), further developed by Mandeville (1970) in an application to the River Ray, Buckinghamshire, and by O'Connell (1970) to the River Brosna, Eire. The first application used three hourly data grouped into events the length of which effectively made it a five day model, and the second application was counted as successful when a predicted value was within $\pm 10\%$ of the observed flow (the residual difference was then set equal to zero). Difficulties were experienced with all three components of the simulation: the model, optimization algorithm and observed data. A considerably modified model was applied by Dickenson and Douglas (1972) to the River Cam, Cambridgeshire, and a second nine parameter model by Blackie (1972) to experimental catchment data from the highlands of Kenya. All this modelling work used the Rosenbrock algorithm to optimize the model parameters, but the real advance was made by Blackie in establishing a realistic and simple approach to the modelling of the routing of effective rainfall through the stores. All subsequent development of the conceptual model at the Institute of Hydrology has followed this basic routing structure whilst concepts have been extended and modified to account for differences of simulation objectives, vegetation and physical structure of the catchments.

A monthly conceptual model was developed for use in South Africa by Pitman (1973), and a daily version by Pitman (1976). The performance of this model was then compared successfully with other models of different complexities, Pitman (1973), available in South Africa.

2.3 Early work on the application of optimization algorithms.

One of the first conceptual models to achieve licensed use world wide was that developed by Crawford and Linsley (1966); the model was known as the Stanford Model. Development work had been in progress since 1960 and it was finally licensed as a modelling package described by Johanson et al (1980). Depending on the configuration used the model had between 20 and 30 parameter values that have to be fitted numerically to any given set of catchment data. Four of these could be optimized automatically in the 1980 version. Its use by inexperienced operators was fraught by difficulties in fitting these parameters and choosing the configuration. This may have hindered the general acceptance of the conceptual model as a useful tool, and subsequent work by others using different models generally encountered the same problems of finding physically significant optimal values for the parameters.

A thesis was presented by Ibbitt (1970) which examined some of the optimization algorithms available, and the results of this work were published in a joint paper by Ibbitt and O'Donnell (1971). The later papers, Ibbitt and O'Donnell (1971), Ibbitt (1972), described in detail some of the problems that catchment models present to automatic fitting methods using real data with artificially introduced errors.

This work was an important contribution to hydrologic modelling in that it appears to be the first research into a choice of optimization algorithm for this type of problem. The algorithms investigated were:-

1. Univariate search.
2. Rotating co-ordinate search.

3. Rotating co-ordinate search with quadratic interpolation.
4. Mutually conjugate direction search.
5. Deflected gradient search.
6. A least squares method with implicit matrix inversion.
7. A least squares method with explicit matrix inversion.
8. A least squares method using Levenberg parameters.
9. A stochastic search method.

To avoid difficulties of robustness caused by the length of record and the form of the objective function, a special form of the latter analogous to the coefficient of variation was devised. This is dimensionless and the effect of the length of record was apparently removed by normalizing the function:-

$$F_{norm} = \frac{\sqrt{\frac{F}{n}}}{\sum \frac{Q_o}{n}} = \frac{\sqrt{nF}}{\sum Q_o} \quad (2.4)$$

Where F_{norm} is the normalized objection function, F is the same least squares function given in equation 1.3, n is the number of runoff ordinates, and Q_o are the observed runoff ordinates. This form of the objective function allows comparison between the degree of fit for sets of data differing in length and variation.

The conclusions of this research were that the Rosenbrock method (Nos. 2 and 3 above) was the best method in obtaining a 'correct' solution irrespective of the cost of obtaining that solution. This had a considerable effect on modelling at that time, and a modified version of this algorithm came into use at the Institute of Hydrology.

In spite of its considerable influence in the field there were three possible shortcomings to the research work. The first problem was that the data used was 'real' but an observed record of only a short batch length could be loaded for each model run. The data came from only one catchment, the Ray experimental catchment operated by the Institute of Hydrology; although field data had been collected in continuous data sets from 1963, the model data arrays did not allow more than a short length to be loaded at a time. Continuous processing using an 'overlay' technique was not included in the programs. All dynamic physically based models are unstable at the beginning of a data run due to errors associated with the initial state of the system; these are usually stabilised over the first two to three months in the case of a hydrological model.

None of the nine parameters to be optimized dealt with the initial state of the system, and whatever the assumptions made about the initial state of the storages this would bias the parameters obtained by optimization over such short periods. The normalised objective function, equation 2.4, would still be sensitive to large fluctuations in n for such small values of m the size of the data subset, and so causing problems with the search for a minimum.

The second shortcoming was the failure to identify the Nelder and Mead (1965) algorithm as one which was in current use, and which should have been included in the comparative tests although nine algorithms were already a large number to test.

The third was that all nine parameters were 'floated' in each run of the algorithm, and that there was no attempt to

establish the need for, or identify, a strategy for the automatic optimization of sub-sets of this number of parameters. Considerable difficulty was experienced with subsequent work at the Institute in trying to optimize all parameters in a large model at the same time. This was probably caused by interdependence of the parameters.

A second attempt at this latter problem was made by Pickup (1978) which tested the efficiency of algorithms AND strategies for automatic calibration of rainfall/runoff models. This was an extension of work by Johnston and Pilgrim (1973) in which the Boughton 12 parameter model using daily data, Boughton (1965), was modified and then used to generate a sequence of flows. Various optimization algorithms were then applied in an attempt to recover the original generating parameters from differing starting values without success. This research objective shows the importance which was then attached to the determination of universal parameters which would relate to their physical significance rather than the model in which they were used.

Pickup modified the Boughton model for use in humid tropical conditions and examined its concepts in detail. From the earlier work it was shown that some of the parameters were interdependent, some were insensitive and that there were discontinuities in the response surface. The least squares equation obtained after minimisation of equation 1.3, was used by both the studies as the objective function.

Four optimization algorithms were used:-

1. The simplex method of Nelder and Mead (1965).
2. The Powell (1965) method of direct search.
3. The Davidon-Fletcher-Powell method (1963).

4. The Rosenbrock algorithm (1960).

To assess the four algorithms a sequence of flows was generated from a synthetic rainfall record which contained both wet and dry periods. This was done to ensure that all the paths in the model were used and that all of the parameters were active at some point in the sequence. Again the parameters were displaced from their original generating values and the algorithms were used to try and recover these original values but without much success. Presumably convergence was achieved to another local optimum. The Nelder and Mead method was the most successful approach to the original parameter set followed by the Powell method. Rosenbrock's algorithm produced reasonable estimates of two parameters, and the Davidon method failed to find any of the 'correct' parameter values. No reason was given for this latter failure by Pickup. If one looks at Chapters 9 and 10 at the end of the author's thesis it is possible that, as the method examined by Pickup depends on the use of a gradient vector, the indeterminacy of the response surface may have been the principal factor.

Tests were made by Pickup on the sensitivity of parameters by varying them by $\pm 50\%$, and examining the changes in model outputs when 'wet' and 'dry' sets of synthetic rainfall data were used. These results were then used to identify a sensitive parameter sub-set affecting model output, and all these parameters were then scaled so that their values lay between 1 and 10. This work was then applied to real data for the year 1970 from the Tua River in New Zealand and obtained lower objective function values for all optimization algorithms tested than those found by simply 'floating' all the 12 parameters. The parameters were unscaled for use in the

model programs called by the algorithms to give the objective function values.

The final conclusions of the paper found that direct search algorithms were more effective than the gradient method which was unable to escape from local optima on the response surface. This work was important in that it considered the three parts of the modelling technique together: the model, the optimization algorithm with parameter scaling, and the aspects of the data sets.

2.4 The St. Lucia Model.

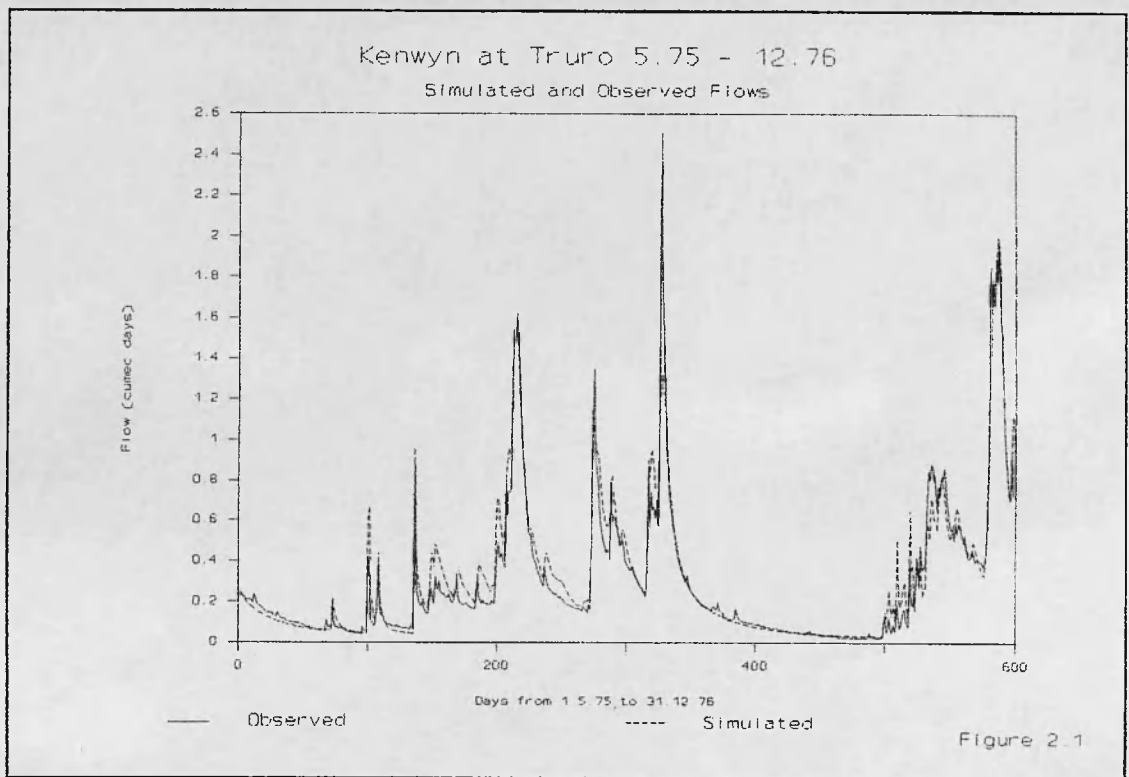
A restricted nine parameter model was evolved for use in research for this thesis from the version of the conceptual model developed by the author which had nineteen parameters available for optimization. The actual model still has nineteen parameters but ten of these were fixed at realistic values obtained from previous modelling projects carried out by the author. A model was required by the UK Overseas Development Administration for use by the Hydrological Department of the Windward Isles for application to the isle of St. Lucia. The model had to be capable of real-time flow forecasting, and a package was developed for use on an IBM PC/AT as detailed in the manual by C.Eeles and Y.Parks (1986). This model was subsequently tested on various catchments by the Applied Hydrology Group at the Institute and its success was such that it was decided it had a more general application than the St. Lucia one. This model became the basis of the modelling package discussed in the next section.

2.5 HYdrological Rainfall/RunOff Model (HYRRROM).

This model has become widely used since it was published, Eeles, Parks and Barr (1989), under its acronym of HYRRROM as a package, and is in general use throughout the world by Universities, United Nations organisations, Polytechnics, Hydrology departments and water resource consultants. For example it has become part of the University of Newcastle civil engineering MSc. course, UNESCO courses in France, and the UNESCO ITC water resources course at Anschede in the Netherlands. It is in use by the UK National Rivers Authority, University College, Ireland, the Alberta Research Council, Canada, the University Politecnica de Catalunya, Spain, the Singapore National University, and the Gyeongang National University, Republic of Korea.

The model has also been the subject of a PhD. thesis at the University of Southampton, I.Son (1990), in which it was used to model the hydrological effects of land use change in small catchments in Zimbabwe and Kenya. The menu interface for the modelling package has now been translated into French and Spanish.

The first success using this very limited version was the modelling of the Kenwyn river at Truro (Eeles 1990); a very mixed land use catchment of 19 km² area over the year of extreme drought and subsequent recharge in 1976, Figure 2.1. This is a very testing period for the use of any hydrologic model. The full data run of input Rainfall and Evaporation data together with observed flow ran from May 1975 to December 1976. The simulation is shown as the dotted line in Figure 2.1 compared with the solid line which shows the observed hydrograph. Both sets of data are in cumec days (average cubic metres per second for 24 hours).



The simulation of the baseflow in both summers is good, and the recovery from the extreme drought of the recession in 1976 is followed well with the exception of three intense rainfall events in September. The peak flow on the 21st. March 1976, just about 320 days, is underestimated by nearly a cumec day. However, there was heavy snow during this period and prior to this peak the flows are underestimated which suggests that snow was accumulating and not forming part of the runoff - then the melt contributed to the very high observed peak. This is a typical simulation response of the model to snow as there is no snowmelt routing component available.

It can also be used as a model using monthly data inputs distributed daily with the usual restrictions applicable to any monthly model: it simulates observed stream flow well when the inputs are changing slowly, but in rapidly changing periods the total volume is a good estimate but not for individual months within the period.

The basic model has five stores representing the vegetation cover, the soil surface layer, soil profile, groundwater store and the surface channel store; outputs from the channel store and the groundwater store are each delayed by a time period which allows them to be combined as flow from the basin, Figure 4.1 Chapter 4. The model is more complicated than it appears from the nine parameters that are available for optimization, but still follows the basic structure and concepts discussed by Blackie and Eeles (1985).

2.6 The Land Use Model.

The original version of this model was developed from the nineteen parameter version mentioned at the beginning of Section 2.4. This land use version was used to assess the effects of afforestation on the water yield from the gathering grounds of the Elan Valley reservoirs in Wales; a study commissioned by the Welsh Water and Severn/Trent Water Authorities, Eeles and Douglas (In preparation). Simulations were made, after calibration of the model, of the resultant flows for seven different levels of afforestation over 53 years. These were then used to estimate the operational and financial consequences of the land use changes. Although no direct comparison was available at the daily interval, the annual totals of evapotranspiration losses relative to potential demand agreed with grassland, heather and forest estimates made from process studies at Plynlimon, Wales and elsewhere - Kirby, Newson and Gilman (1991) and Calder (1986).

The loss functions for transpiration and interception normally used in the HYRRROM model algorithm were applied to the grasslands, and models developed by Calder and Newson (1979)

were used for losses from coniferous forest and heather. Evaporation losses from the open water areas of the reservoirs were estimated using the Penman open water potential evaporation index, E_0 .

A more direct comparison was made when the model was calibrated on the gradual clear felling of the partially forested Hore sub-catchment, area 3.17 km^2 , of the Plynlimon experimental catchments, Gross et al (1989). The model was shown to underestimate systematically low flows during the felling, although relative changes in the flow regime were satisfactorily modelled. At a later date it was found that the datum for the gauging structure was in error by 9 mm due to modifications to the flume at the beginning of the felling period; this accounted for the apparent bias in low flows estimation. The bias being most apparent at the low flow end of the flume calibration, and virtually unnoticeable at high flows.

CHAPTER 3.

Application of conceptual models to hydrological problems and stream flow forecasting.

3.1 Objectives of catchment simulations.

These objectives fall into two main categories: the study of the hydrologic cycle and of the effect of man on the natural environment. The general objectives have been discussed in section 1.4 but consideration of the outputs from a good physically based conceptual model and their application is an important exercise in this type of modelling.

The possible outputs from a conceptual model vary according to the complexity of the processes simulated by the model structure. However, even the simplest version of this type of model should be capable of the following outputs for given precipitation data:-

- a. Streamflow estimates.
- b. Evapotranspiration loss estimates.
- c. Soil moisture deficits and storage changes.
- d. Changes in groundwater storage levels.

As a direct consequence of the general use of only (a) in the objective function with which the model is calibrated, any errors and bias due to the conceptual structure of the model may reduce the accuracy of the estimates of the other three. To take one example: the absolute soil moisture storage values may be biased but the changes in storage are probably predicted with reasonable accuracy. This is also a stricture which applies to the sampling and processing of soil samples, but in any comparison between model output and the field

samples it has to be remembered that the latter refer to a point in the area while the model output refers to the whole of the general area.

Because of its position in the 'top down' structure of the model the groundwater storage level estimates tend to have the greatest bias and inaccuracy, and because of this an analysis of the streamflow recession curve should be made, if possible, to establish field values of the relevant parameters. It is sometimes impossible to do this due to the complications in the hydrograph from timing delays and to the distribution of inputs.

The uses to which the simulated outputs are put fall broadly into the following categories:-

1. Quality control and infilling of missing data.
2. Extension of historic flow records.
3. The generation of synthetic data series for civil engineering design work and other applications.
4. The assessment of water resources.
5. Water resources management including real-time forecasting.

The applications to which these are put fall into the two main categories given at the beginning of this chapter. A more detailed examination of the general methods of approach are made in the following sections using results from model applications.

3.2 Quality control and infilling of time series data.

Where hydrological records have been collected over a long period of time from research or monitoring programs these generally become the basis for a wide range of modelling and

statistical analyses. To ensure that the value of results derived from such data bases are not biased by discontinuities, trends or other errors arising from observational or processing methods, a variety of quality control and infilling techniques are used by hydrologists and engineers.

Regression, double mass plotting and time series analysis are perhaps the most commonly used techniques, and are very useful in analyzing the structure of data. These can also be used in verifying the simulations made by models, which in turn can be used to identify the periods over which observed data errors are present and help to indicate the sources of error. These can be due to processing, to unsuitable positioning or unsuitable exposure in the field, and to instrument malfunction.

The use of a model has much to commend it when used for infilling and data extension in that the simulation of the physical processes transforming precipitation inputs to outputs contains, directly through the storages, a persistent memory of antecedent conditions. The basic approach in this type of application is to calibrate the model on the period of record considered initially to be the most reliable. If a good correlation with the observed data is achieved then this in itself is some verification of the initial assumptions in the design of the model. Any errors in this section of the data run are likely to be systematic and consistent. The model is then verified by its simulation of the rest of the data run. The size and duration of discontinuities can be highlighted by plotting the cumulative time series of residuals. Sudden discontinuities in this plot arise either from errors in the

input or observed flow data, or from errors in the model assumptions and design!

3.3 Record extension and synthetic record generation.

The Balquhiddar, Scotland, study of upland water resources provides an example of the use of catchment data to extend a data series so that a statistical study can be made of the longer record. The paired catchments were set up in 1983 giving seven years of data at the time of the study, but at least 25 years of data were needed to allow the statistical results to be extrapolated to other upland areas for the paper by Gustard and Wesselink (1993). These results are to be used to form a basis for the design of dams and reservoirs, and set limits on the afforestation of the areas forming gathering grounds.

The observed rainfall record starts with the instrumentation of the catchments in 1983, together with the Penman E_f data for the evapotranspiration loss functions. This E_f data is processed from meteorological observations from the Kirkton High and Upper Monachyle automatic weather stations for each catchment. The model algorithm requires input data of rainfall and E_f and so a means of extending the historic record back to the beginning of 1964 was required to give 25 years of historic record.

The use of the mean annual E_f data set has been shown to work remarkably well with simulations of soil moisture changes for grassland in Calder, Harding and Rosier (1983), and subsequently for pine forest and heather by Hall (1987). It obviously cannot be used to simulate conditions of severe drought when evaporative potential demand is at its greatest,

or those in flood conditions when the potential is at a minimum. However, the mean annual data set has been accepted and used in extending the historic record for other studies, and so it was decided to use it as there was no immediately available data from a suitable meteorological station covering the period back to 1964.

The search for a suitable raingauge record was made in the UK Daily Rainfall Archive at the Institute of Hydrology using the Great Britain National Grid within the rectangle defined by 2300, 7100 and 2700, 7400. The only suitable continuous record was that from the Lochay Power Station (2546, 7350) at an altitude of 116 m whilst the two catchments are above 300 m rising to nearly 2,000 m.

The data from this station was compared by regression analysis with the rainfall data from Balquhiddar 1983-1988. The best fit was obtained by a simple linear regression passing through the origin for both sets of data. The results were:-

$$R_{\text{Monachyle}} = 1.33 \cdot R_{\text{Lochay}} \quad (3.1)$$

where $R_{\text{Monachyle}}$ and R_{Lochay} are the rainfall data.

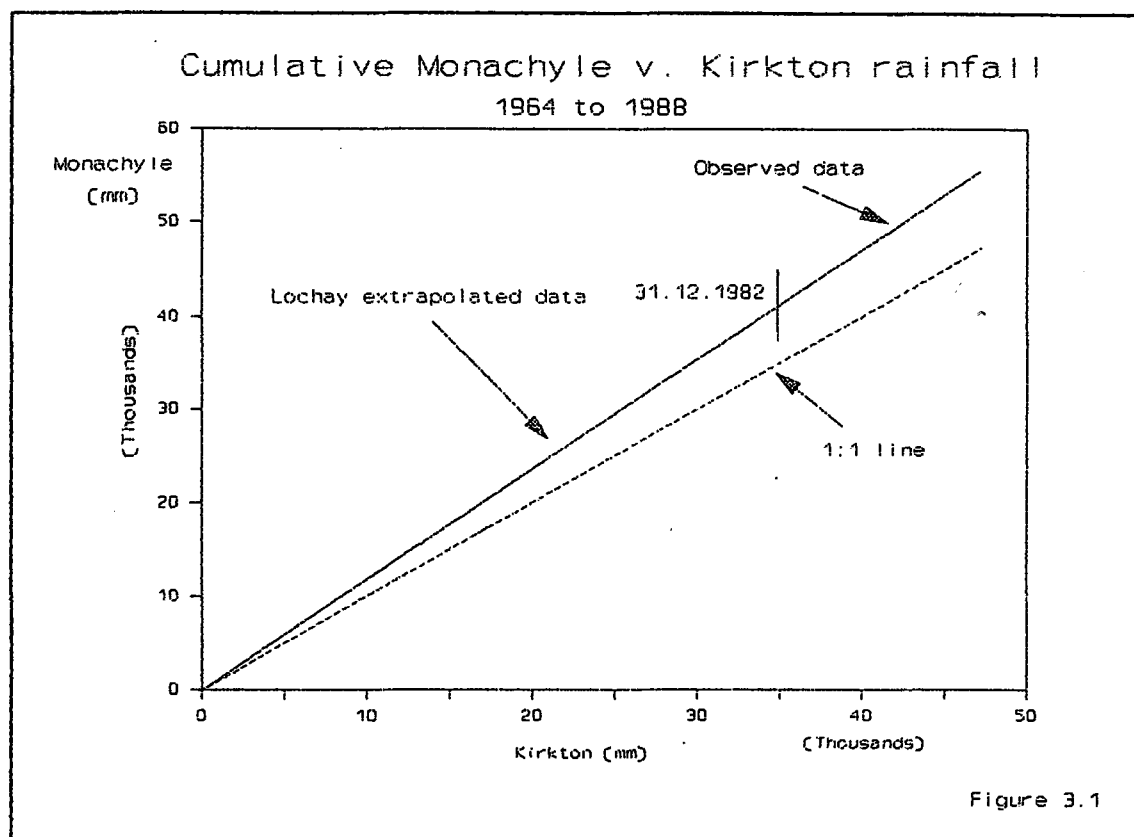
The Monachyle standard error of estimate for the regression coefficient is 0.012 and regression correlation is 0.89.

$$R_{\text{Kirkton}} = 1.13 \cdot R_{\text{Lochay}} \quad (3.2)$$

where R_{Kirkton} is the rainfall data for the Kirkton.

The Kirkton standard error of estimate for the regression coefficient is 0.010 and regression correlation is 0.89.

The difference of nearly 18% between the coefficients stems from the higher rainfall in the more westerly of the two catchments, the Monachyle.



The graphs comparing the two sets of generated rainfall data from January 1964 to December 1982 cumulatively, Eeles and Blackie (1993), are given in Figure 3.1. The upper end of the lines from January 1983 uses the cumulative observed data from each catchment.

The land use model calibrated and verified on the run of observed flows, Eeles and Blackie (1993), was then used to generate 25 years of simulated streamflow. This is the minimum time period for such curves as the 'flow duration curve' to make valid stochastic predictions. The normal starting point in designing a reservoir, a water supply abstraction offtake or a bridge is an analysis of the existing streamflow record.

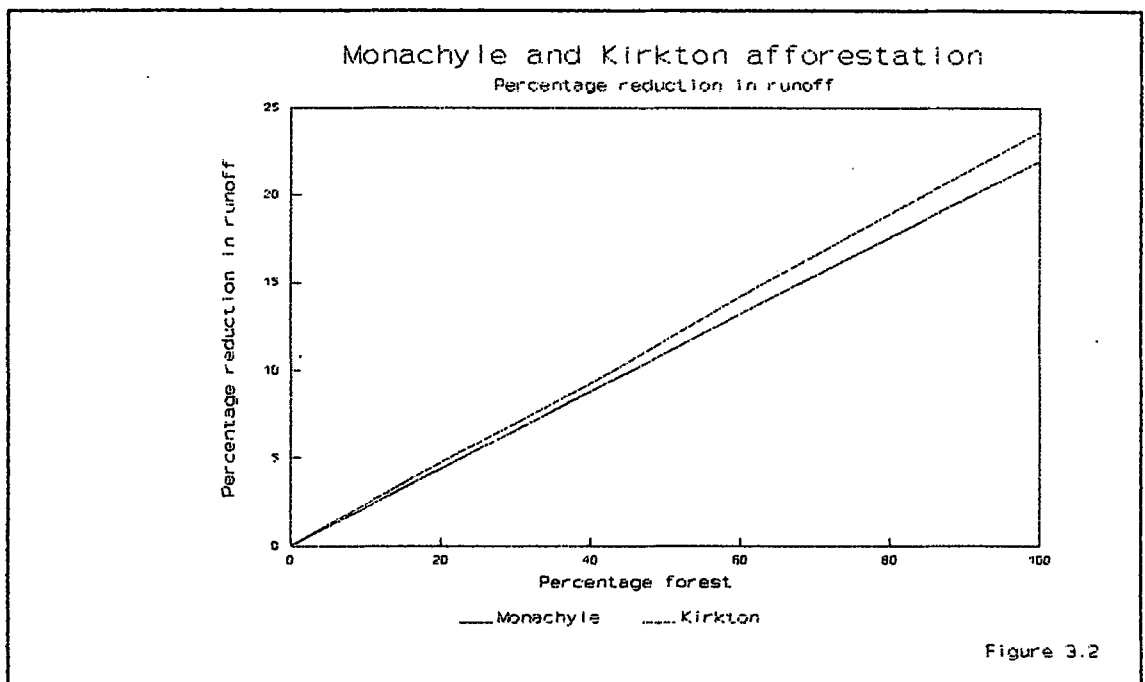
From the flow duration curve and the statistics of the flow extremes the necessary design parameters are obtained. Unfortunately, the flow records are usually too short to make the range of uncertainty acceptable for the purpose of the design. To reduce this uncertainty, time series methods are normally employed first to determine the structure of the existing record, and then to generate sequences of synthetic data which are statistically indistinguishable from the original record.

This approach relies on the historic record being long enough to provide an adequate definition of its structure. If it is not, then a suitable conceptual model can be used to extend it as far as rainfall and evaporation records exist, and then rainfall data from a distant station together with the annual mean evaporation record can be used to allow generation of a synthetic record.

3.4 Water resources assessment.

In the previous section the use of parametric models to extend records for design purposes was discussed. This has obvious applications also in the assessment of water resources, provided no major changes in land use occur within the period of interest. In using conceptual models, or any other models which are calibrated on a historic record, for record extension it is assumed implicitly that the catchment response characteristics do not change. If major changes in land use, in drainage or in river control systems occur this assumption becomes invalid unless the model has been designed to allow for these changes. Extensive drainage will alter the runoff response and river training works will modify the timing and

shape of flood hydrographs. However, a conceptual model has been designed to allow for changes in land use. The model used in the previous section was calibrated for the Kirkton on the period May 1984 to December 1985 during which no changes took place, and then verified successfully on the period January 1986 to December 1988 during which areas of the Kirkton were being clear felled. The clear felling modified the evapotranspiration rate from the catchment and hence the streamflow volume, and this change was simulated well by the model. Different versions of this model have been successfully tested in other areas, e.g. Gross et al (1989).



The results from these extended data sets, when the model has processed them for different levels of afforestation, show how they reduce the runoff from a catchment. The results of the simulations for the Monachyle catchment are shown in Figure 3.2, with the present proportion of grasslands to heather moorlands kept at 50.4:49.6 for different areas of forest. The ratio of simulated actual evapotranspiration to observed E_T

ranges from 1.053 without forest to 1.737 completely forested. The total area of forest is reduced by 25% to allow for rides, roads, etc.

Figure 3.2, also gives the simulation results for the Kirkton with its higher present proportion of grasslands to heather moorlands of 75.6:24.4 for the different areas of forest. The ratio of simulated evapotranspiration to observed E_T is from 0.761 without forest to 1.425 with 100% forest.

The two afforestation graphs shown in Figure 3.2 have the qualification that they represent afforestation and not deforestation. They can only be considered in the latter case if the cleared areas are recolonised in the original proportions of grassland to heather. The difference in the reduction of flows is assumed to be a result of the higher proportion of heather to grass in the Monachyle and the higher rainfall.

This sort of restriction limits the detailed use of such models but as a general guide they are useful in the absence of observed data. Whilst it is possible to use a parametric conceptual model to predict these effects, or to estimate flows from an ungauged catchment, the credibility of the results will be low unless independent checks can be made.

3.5 Management of water resources.

The use of parametric conceptual models now has an accepted place within the assessment and design stages of water engineering schemes, but their use as a management tool is less well established. In comparing the long and medium term effects of possible water resource management strategies they can be very efficient as shown by Eeles and Douglas (In

preparation). In this study the original version of the land use model was used to generate synthetic data runs from 1931 to 1984, which were then used as inputs to the water authorities operational programs to assess the financial effects of changes to the gathering grounds of the Elan Valley reservoirs in Wales. Limits were set on the levels of afforestation which could be sustained using current water resources. It was predicted that, water would have to be abstracted from new bore holes, new water processing works would have to be constructed, and abstractions taken from more expensive water resources if further afforestation was allowed.

3.6 Future developments in use and extension of parametric conceptual models.

Just as the development of the parametric conceptual model was dependent on the existence and availability of the mainframe computer, so new applications of these models depend on the application of Personal Computers and microprocessors controlling data capture by instruments. The capture and processing of raw data and its telemetry back to the main data base has become a viable system for environmental experiments throughout the world. The data is either stored on a microprocessor controlled logger and transmitted via satellite, or directly to the base station through 'slave' transmitting stations. With this ability to use distributed processing at the logger stage to simplify the telemetry requirements a virtually realtime reporting network can be developed using cheap computing hardware.

The accuracy of a good conceptual model matches that of the large distributed model without the disadvantages of the

latter. The computing requirements of this distributed type of very large model are such that they rapidly approach the resource limits of mainframe computers, and the only hope for further advance in their application lies in the application of parallel processors or the development of more powerful computers. Attempts have been made to mount this type of model on a personal computer, but a successful package does not seem to have been developed and published. There are possible future developments in the combination of the conceptual and distributed models, with the parametric model concepts streamlining areas of application to basins of the mechanistic models.

Parametric models are already in use by National Rivers Authority Severn/Trent division, Dobson 1993, for flood forecasting in the Severn and Trent basins. This system aims to achieve a four hour forecast on major rivers of their rising above three warning levels. Two models are in use: a twenty six parameter model with ten controlling the major part of the rainfall/runoff response, and a second model used for flow routing. Data capture is by instruments mounting their data on loggers, and the system interrogated by use of landlines with a overall time of sixteen minutes per basin. The model execution time is three minutes for each basin with automatic alarm raising and forecast generation. The models are applied to sixty five sub-catchments in the Severn and forty eight in the Trent, and parameters are derived by optimization for the 40% of catchments, which are gauged. For the ungauged catchments parameters are estimated by regionalisation based on the regression of model parameters against measurable catchment characteristics. Improvements to this system have concentrated

on data collection, and the next step is the recalibration and improvement of the models. The potential of such systems is considerable with the improvement of Meteorological Office rainfall forecasts and the possibilities offered by distributed processing on small and more powerful computers; the present system is based on a DEC MicroVAX.

Future applications of HYRRROM depend on making the modelling package more flexible, and on extending the optimization part of the package to a hybrid algorithm such as the one developed for this research. With the widespread use of the package on hydrologic problems it is hoped that 'feedback' from users will further improve the package, and show extensions to the range of use.

The Land Use model developed during this research needs refinement by further experimental process studies for different vegetation types to extend the range of 'loss' functions, and for the collection of different data sets from basins where the timing and area of land use change is recorded.

CHAPTER 4.

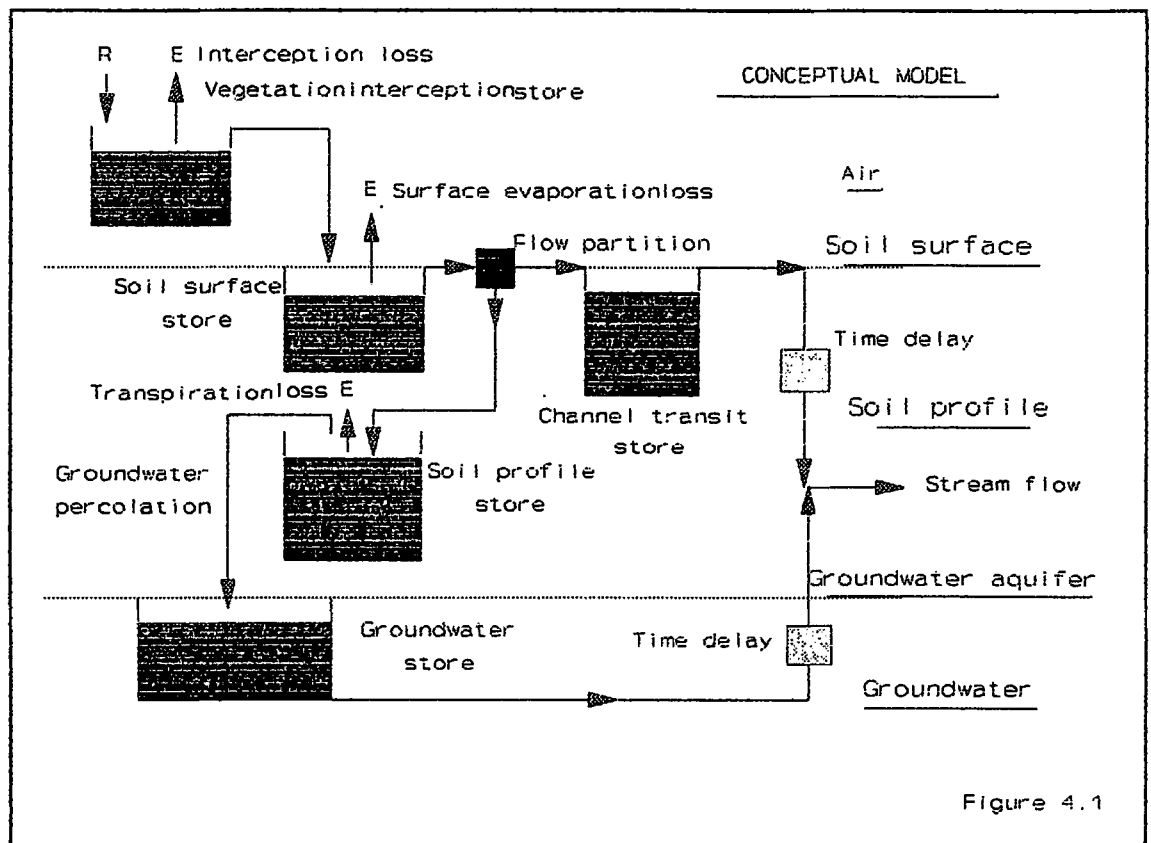
The parametric conceptual models.

4.1 The generalised model concepts.

The general concepts of this type of model, used to simulate the rainfall/runoff regime from catchments, are as old as the first hydrologic models to be run on a mainframe computer. In some models the structure is implied by the ordering of the model equations. Since these equations reflect the order of processes occurring in the natural structure of a catchment, the only conceptual variation is in the detail. For example, the movement of moisture can be considered as a simple vertical movement within a unit time. Alternatively interflow, 'horizontal' flow at different interfaces in the model profile, can be considered as a flow path into the stream at intermittent times to simulate the finer detail of the hydrograph.

The simple concepts are shown in Figure 4.1 with the model profile divided into three sections by the air/soil interface and soil/groundwater interface. Above the air/soil level is the vegetation interception store which receives the precipitation input, R , and from which the interception and evaporation loss, E , takes place.

Further evaporation losses occur from the soil surface store (the humic or litter layer). The throughfall from the vegetation canopy is input to this store. The store fills and when it is full 'rapid' response runoff occurs together with infiltration into the main soil profile store at the flow partition. This rapid response runoff adds to the transit store which represents the catchment dendritic channel system.



The infiltration is added to the soil profile store, and losses occur from this store in the form of vegetation transpiration extracted by the roots from the soil, together with percolation through the soil to the groundwater store. The 'slow' response flow from this groundwater store has a time delay imposed on its component of the streamflow from the outfall of the catchment. A shorter time delay is applied to the rapid response component of the streamflow to represent time of travel through the surface channels.

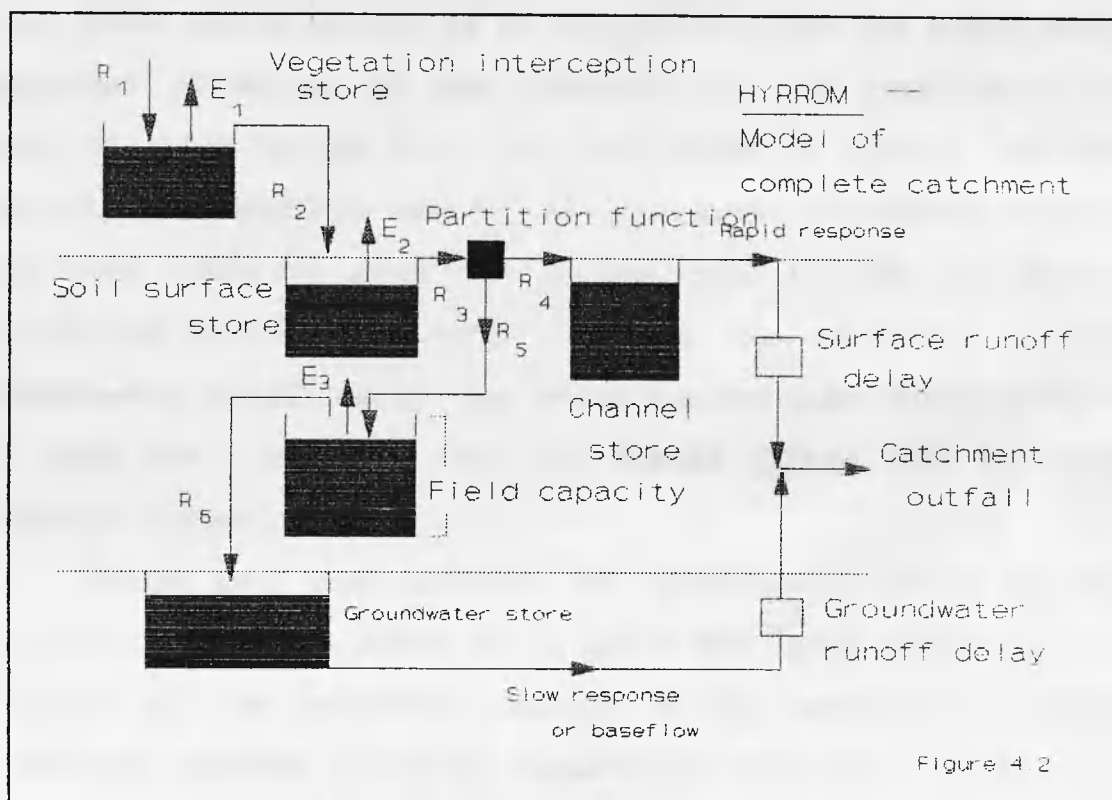
The ordering of concepts in Figure 4.1 shows the sequence of processes which controls the dynamic 'top down' operation of the functions in the model structure on the observed data input of precipitation, R . These rainfall data inputs are reduced at each stage, or module, to 'effective' rainfall by the evapotranspiration loss functions, E . In the model developed for use in this research, which later became

the basis of the model package HYRROM, the E-functions are designed to use the potential Penman open water evaporation variable, E_0 , as an index of evaporative demand; the value of E_0 is derived by Penman (1948) from meteorological data as shown in Appendix I. This 'loss' from the system reduces the potential demand at each stage. At the end of each time interval the residual potential demand is set to zero, and a new value is input for the next time interval.

4.2. The HYdrological Rainfall/RunOff Model (HYRROM).

The HYRROM model is described first since the functions and parameters used in HYRROM to represent the whole catchment form a sub-set of the main land use model. This land use model is more complicated as it models the grassland areas as well as two other areas of vegetation using the appropriate Calder (1986) equations. The Calder equations are empirical equations fitted by optimization of the parameters using observed data from vegetation process studies on stands of pine forest and experimental plots of heather to estimate interception and transpiration losses. Grassland areas are modelled using the original vegetation catchment functions from HYRROM. The 'loss' functions in particular from HYRROM are used to model the grassland area losses. The general outline of HYRROM is shown in the schematic diagram Figure 4.2, and the land use model in Figures 4.3 and 4.4.

The first store in Figure 4.2 represents the vegetation interception store: the observed rainfall, R_1 , is first added to the store, and then the interception loss generated by the potential evaporation demand, E_1 , is taken from the store at the following estimated rate.



The contents of the interception store, $S_{intercept}$, remaining after any overflow are depleted at a rate for the i -th time interval by the evaporation loss function

$$E_1 = \frac{\Delta S_{intercept}}{\Delta t_i} = p_2 \cdot E_{obs} \quad (4.1)$$

where p_2 is the Penman evaporation factor and E_{obs} is the observed Penman open water evaporation for the time interval.

The total water lost to evaporation cannot exceed the amount held by the interception store. The parameter p_2 is related to p_5 , the vegetation soil transpiration factor, in order to reduce the number of parameters in HYRRM by the empirical equation derived from previous modelling work:-

$$p_2 = 1.0 + p_5 \quad (4.2)$$

This first store concept is an abstraction from the complicated physical situation of the interception and evaporation of precipitation by the leaf area and stems of plants, and the detention of surface water. All the model processes take up one time interval (one day in the case of the two models developed during this work). The one exception is the groundwater runoff delay, p_{23} , which has one time interval added to mark the transition from the closed system into the open channel system.

During each time interval the interception store has the rainfall input, R_1 , added to it until the model state initial content p_3 , and subsequent values of the interception store contents, reaches its store capacity p_1 when the overflow, or throughfall, R_2 , is the input to the soil surface store. The ongoing 'effective' rainfall, R_2 , is calculated as

$$R_2 = \begin{cases} 0 & \text{when } (p_3 + R_1) \leq E_1 + p_1 \\ (p_3 + R_1) - E_1 - p_1 & \text{otherwise} \end{cases} \quad (4.3)$$

where p_3 is the initial contents of the interception store at time t_0 and subsequently the contents of this store. p_1 is the size of the interception store.

The initial contents p_3 at time $t = 0$ are shown as a parameter to be optimized in Table 4.1 for the Land Use Model. This is done as the total volume predicted by the simulation can be biased by an incorrect value. This store must fill and then empty at the right time to allow a good subsequent fit to the hydrograph. At each time interval the state value of the surface store, p_6 , is added to the effective rainfall, R_2 , and the effective rainfall passing beyond the store when it obtained in previous model simulations.

overflows, R_3 , is calculated by subtracting the size of the surface store, p_4 . At the end of the i -th time interval the value of the soil surface store contents changes as

$$R_3 = \begin{cases} 0 & \text{if } p_6 + R_2 \leq p_4 \\ (p_6 + R_2) - p_4 & \text{otherwise} \end{cases} \quad (4.4)$$

In HYRRROM the model initial state contents of the interception store and the surface store are set to zero. This dry condition also applies to the water in transit across the catchment surface; the channel store contents are again set to zero. The other initial state parameter of the soil profile store (the soil moisture deficit), p_{10} , is set to 10 mm; this appears to be a good average initial deficit state for this soil profile store in the dry conditions early in the year in the UK.

Since the humic and litter layer represented by the shallow soil surface store is very much less dense than the lower soil profile, the loss process is simplified to one of straightforward evaporation; E_2 is estimated in a similar fashion to E_1 in equation 4.1 as

$$E_2 = p_5 \cdot (E_{\text{observed}} - E_1) \quad (4.5)$$

The initial soil surface store contents p_6 together with R_2 again overflow the size of the store, p_4 , and the resultant overflow, R_3 , is partitioned between surface flow and infiltration into the soil profile. The size of the soil surface store, parameter p_4 , is fixed at 9.5 mm for use in HYRRROM. This figure was arrived at from an average value obtained in previous model simulations.

$$R_3 = \begin{cases} 0 & \text{If } R_2 + P_6 \leq E_2 + P_4 \\ R_2 + P_6 - E_2 - P_4 & \text{otherwise} \end{cases} \quad (4.6)$$

$$\begin{aligned} R_4 &= K_i \cdot R_3 \\ R_5 &= (1 - K_i) \cdot R_3 \end{aligned} \quad (4.7)$$

Where i is the i -th. interval and K_i for that interval is

$$K_i = P_7 \cdot [e^{P_8 \cdot S_{smd}} + e^{P_9 \cdot R_3} - 1] \quad (4.8)$$

The initial soil moisture deficit, p_{18} , is used for the initial time interval, t_1 , and then replaced by the new soil moisture deficit, S_{smd} , estimated by the model for time t_2 and succeeding times.

$$(S_{smd})_{t_i} = (S_{smd})_{t_{i-1}} + R_5 \quad (4.9)$$

The effective rainfall infiltrating into the soil, R_5 , is the residue from the rapid response component (equation 4.7).

The rapid response runoff, R_4 , is added to the channel store contents, S_{chan} , and released from this:-

$$Q_s = \frac{\Delta S_{chan}}{\Delta t_i} = P_{11} \cdot S_{chan}^{P_{10}} \quad (4.10)$$

This release, Q_s , is delayed by the time interval p_{12} when it forms part of the catchment yield at time $(t_i + p_{12})$. p_{11} is the channel routing store contents factor.

The infiltration component, R_5 , is added to the soil profile store, S_{soil} , and when the soil moisture deficit is less

than zero (excess moisture in the store) the percolate to groundwater, R_6 , becomes:-

$$R_6 = \begin{cases} 0 & \text{if } S_{smd} \geq 0 \\ p_{17} \cdot S_{smd} & \text{otherwise} \end{cases} \quad (4.11)$$

where p_{17} is the percolation to groundwater factor, and S_{smd} the soil moisture deficit.

When S_{soil} has an excess of moisture which cannot be held by surface tension in the soil matrix against the force of gravity i.e. field capacity is exceeded, then S_{smd} , the soil moisture deficit, is negative to represent this surplus - hence the negative sign in the above equation. In the HYRRROM code the slope of this equation, p_{17} , is set to -0.5 empirically to reduce the number of optimised parameters.

After this percolate has been transferred from the soil store then the transpiration, E_3 , is estimated by:-

$$E_3 = K_{reduct} \cdot (E_1 - E_2) \quad (4.12)$$

The full transpiration varies from the Penman root constant, where the vegetation is first under moisture stress, to the point at which it can no longer extract moisture from the soil and it wilts and dies. This process is represented by a smooth cosine curve to show the transition, but there appears to be no experimental data to justify this hypothesis. K_{reduct} , the cosine reduction factor, is given by

$$K_{reduct} = p_5 \cdot \frac{\cos(\theta_r) + 1}{2} \quad (4.13)$$

Here p_5 is the Penman transpiration factor and θ_r is given by

$$\theta_r = \begin{cases} 1 & \text{for } S_{\text{sm}} \leq 0 \\ \frac{S_{\text{sm}}}{p_{15}} & \text{for } 0 < S_{\text{sm}} < \Pi \cdot p_{15} \\ 0 & \text{otherwise} \end{cases} \quad (4.14)$$

The transpiration is limited by a high soil moisture deficit with the maximum value p_{15} and the Penman root constant, p_{16} is set at zero soil moisture deficit in HYRRROM.

The soil percolate is added to the groundwater store, G_{store} , and is then released as the 'slow' response, Q_g ; at the rate ΔG for the time interval Δt_i :-

$$Q_g = \frac{\Delta G_{\text{store}}}{\Delta t_i} = \left(\frac{G_{\text{store}}}{p_{21}} \right) p_{22} \quad (4.15)$$

This is then delayed by p_{23} and released at time $(t_i + p_{23})$ in addition to the 'rapid' response component for the same interval, and so simulates the total river flow Q_{catch} .

The general absence of groundwater data and the problem of extrapolating point readings from wells or piezometer tubes to the whole area of a catchment makes it necessary to set the initial contents of the groundwater store by using the first three values of the observed flow in the starting month, and inverting the model groundwater release equation 4.15. The groundwater delay is first split into its integer and decimal parts to allow distribution between the intervals at the end of the delay

$$p_{23} = (p_{23})_{\text{integer}} + (p_{23})_{\text{decimal}} \quad (4.16)$$

$$G_{\text{tran}} = [1 - (p_{23})_{\text{decimal}}] \cdot Q_{(p_{23})_{\text{integer}}} + (p_{23})_{\text{decimal}} \cdot Q_{(p_{23})_{\text{integer}}+1}$$

where G_{tran} is the output from the groundwater store.

The constraint generated by this inversion is that the simulation must be started when the river stage is in recession and observed rainfall input zero in the intervals immediately before t_1 :-

$$Q_{-2} > Q_{-1} > Q_0 \quad (4.17)$$

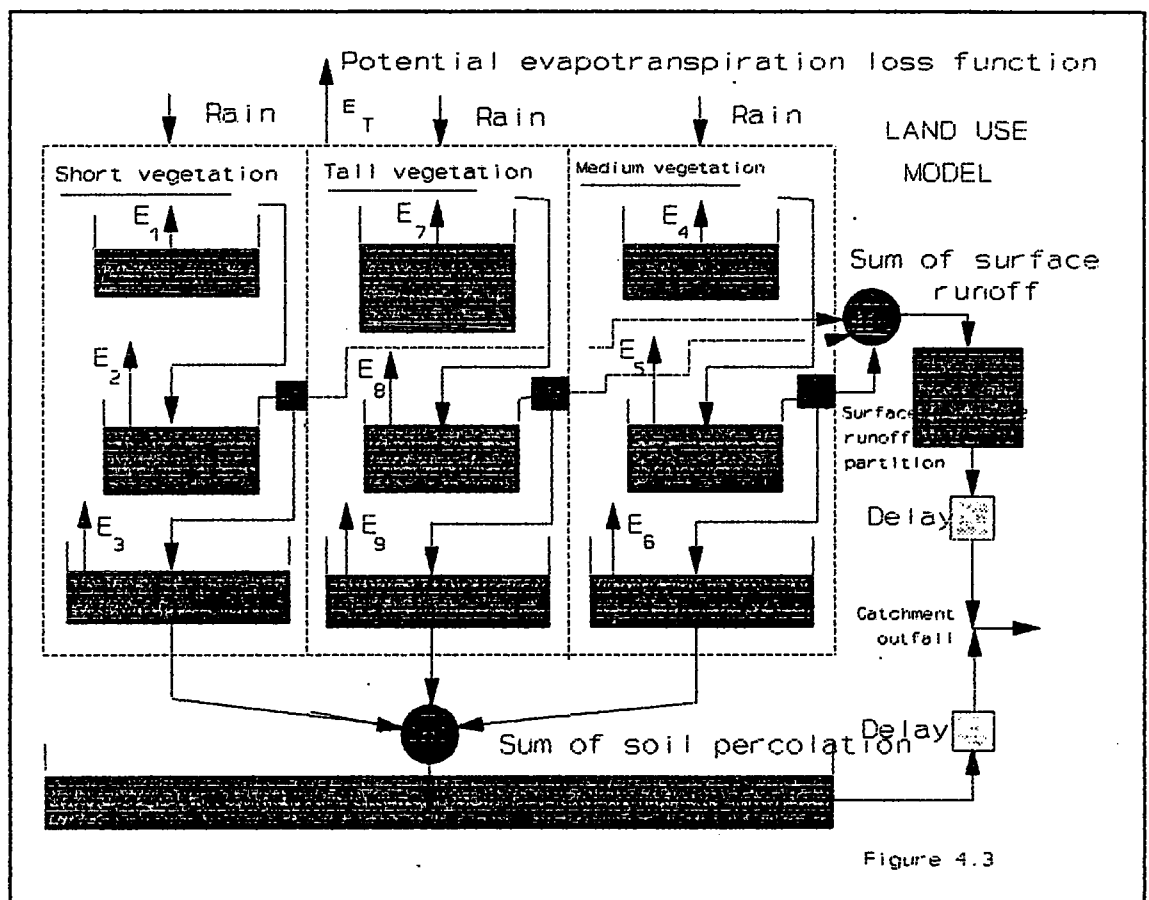
Where Q_{-2} is the observed flow at time t_{-2} , Q_{-1} is the flow at t_{-1} , and Q_0 is the flow at t_0 .

The FORTRAN77 code for the HYRRROM model is given in Appendix 2. The model code as shown works by processing a month of daily data at a time, and then overlaying this data with the next month of data from a file or data stream. This allows the model to accept time series data set of any length, but it was found to be too slow for the I/O of a PC during the operation of the Rosenbrock algorithm, and the length of data is therefore restricted to 10 years in optimisation mode and 50 years in direct simulation mode as the data is read directly into data arrays for speed of access. This in turn causes problems from the large size of the arrays if the random access memory has too many package programs loaded with dedicated areas of RAM memory.

4.3 Structure and concepts of the Land Use Model.

The land use model was developed to provide a routing and storage structure to enable the Calder and Newson (1979) interception and transpiration loss concepts to be incorporated in a model operating on a daily basis. The loss equations, Calder (1986), are based on process studies for coniferous

forest, heather and grassland. The equations were used originally for annual estimates of water use, and only recently has an attempt been made by Hall and Harding (1993) to estimate this on a shorter time base over a year with limited success due to the absence of routing and storage functions. The model was first applied to the Elan Valley reservoirs study, Eeles, Farquharson and Harding (1986), in which changes in afforestation were simulated for five scenarios from 1932 to 1985, to enable its effects on the management of water resources and the financial implications of drought conditions



to be assessed. A separate sub-model is established to represent each different land use as shown in the model general structure diagram Figure 4.3.

To each of these models the daily estimation of rainfall and potential evapotranspiration for the whole catchment are

applied as inputs, and the simulated outputs to the surface channel store and groundwater store are weighted by the fractional area of the vegetation. This process is illustrated by the module diagram in Figure 4.4. In the grassland sector a simple formulation of a factor, p_5 , times the Penman evapotranspiration index, E_T , is used, with a reduction term for transpiration affected by soil moisture deficit. As the deficit increases beyond the Penman root constant, p_{16} , the factor modifying the Penman index is reduced gradually to zero using a cosine reduction function. These are of the same form as equations 4.13 and 4.14 but with a different expression for θ_r to allow a drier range of soil moisture deficit to be used:-

$$K_{reduct} = p_5 \cdot \frac{\cos(\theta_r) + 1}{2} \quad (4.18)$$

where p_5 is the Penman vegetation transpiration factor and θ_r is now given by

$$\theta_r = \begin{cases} 1 & \text{for } S_{smd} \leq p_{15} \\ \frac{S_{smd}}{p_{16} - p_{15}} & \text{for } \Pi \cdot p_{15} \leq S_{smd} \leq \Pi \cdot p_{16} \\ 0 & \text{otherwise} \end{cases} \quad (4.19)$$

This simulates the reduction of transpiration to zero at plant wilting point, p_{16} , under the action of vegetation moisture stress caused by the roots being gradually unable to exert the suction with increasing SMD (soil moisture deficit) to extract moisture from the soil matrix. This starts from p_{15} where SMD begins to cause plant stress to p_{16} where the plant wilts and dies. The maximum SMD for forest is p_{19} , and for heather p_{20} .

The main difference between this model and HYRRROM, apart from the land use structure in three parallel modules, is in the equations used to estimate vegetation interception, evaporation and transpiration losses.

These evapotranspiration models are described below and are designed to use annual data, but have been extrapolated to use simple daily observations of rainfall and the more complicated meteorological data combined in the E_T modification of the Penman index in Appendix 1, to estimate total water use by the vegetation. Their structure and parameters are derived from the results of process studies mounted by the Institute of Hydrology on experimental plots on the Plynlimon Catchments in Wales, at Thetford Forest in Norfolk and at various sites in Scotland. The models have been derived from observations of soil moisture changes, vegetation interception and experimental plot runoff, and by comparison with catchment water balance studies on an annual basis. A more detailed discussion of the models is given by Calder (1986). The heather moorland and forest models estimate the losses due to rainfall interception, the transpiration when the canopy is dry, and the period in each day when the canopy is dry; transpiration is then assumed to be taking place instead of interception losses.

The canopy interception loss, $E_{intercept}$, is calculated from the initial total rainfall, R_{total} , using the following exponential expression:-

$$E_{intercept} = \gamma (1 - e^{-\delta R_{total}}) \quad (4.20)$$

where γ and δ are constants which vary according to the type of vegetation:-

$$\gamma = p_{26} \text{ for forest; } p_{29} \text{ for heather}$$

$$\delta = p_{27} \text{ for forest; } p_{30} \text{ for heather}$$

The values of these experimental constants have been fitted to the observed results of experimental plot interception experiments in Calder (1986) by optimization, and these values are in turn used as a starting point for the optimization process used with the model. Since the model refers to the whole of the vegetation area and the evapotranspiration models are based on small areas of plot experiments, only a limited correlation can be expected between the two. As the models are abstractions of reality there is a degree of 'fuzziness', or uncertainty, about the parameters which can only be reduced by their optimization after the rest of the land use model parameters have been set.

This expression for interception losses gives a daily interception limit for light rainfall but causes the daily overall losses to approach a maximum limit for larger rain storms when canopy throughfall becomes the dominant process. The fraction of the time when the canopy is wet is derived from:-

$$W = \begin{cases} 1 & \text{if } R_{total} \geq 0.99/p_{24} \\ p_{24} \cdot R_{total} & \text{otherwise} \end{cases} \quad (4.21)$$

where p_{24} is a factor to be optimized and $W = 1$ when R_{total} exceeds $0.99/p_{24}$ (mm) the precipitation interception limit from plot experiments. The constant 0.99 is derived from work

reported by Calder (1986) on the daily rainfall and the mean climatological rainfall intensity in the UK. It is not explicitly stated or derived in the work but is implicit, and is also found to fit the new interception limit given for heather in Hall and Harding (1993).

The vegetation transpiration, E_{tran} , is estimated from the following expression when effective rainfall, $R_{through}$, or throughfall is added to the soil surface store and the residue from surface runoff infiltrates:-

$$E_{tran} = \beta \cdot E_T(1 - W) \quad (4.22)$$

where $\beta = p_{25}$ for forest, p_{28} for heather.

Here β is a constant to be optimized from the initial starting value determined by field observations, Calder (1986), and E_T is the Penman potential evapotranspiration. The combined equation consisting of those for evapotranspiration and interception losses for both vegetation types follows the formulation given by Calder and Newson (1979).

4.4 The complete land use model.

The full integrated land use model is shown in diagrammatic form in Figure 4.3 with sections partitioned according to the three land use types: grass or 'flat vegetation', forest or 'tall vegetation', and heather or 'medium vegetation'; the choice of the new names used for the redefined sections is described by Roberts et al, 1993. These more general terms are due to the mixed nature of the vegetation at Balquhiddy in Scotland. For each land use

division of the model the modular structures are processed separately until the outputs are each weighted by one of the n land use areal fractions, A_n , and summed for the total area. These modular structures are as shown in Figure 4.4 with their outputs combined as inputs to the catchment surface response store and groundwater store.

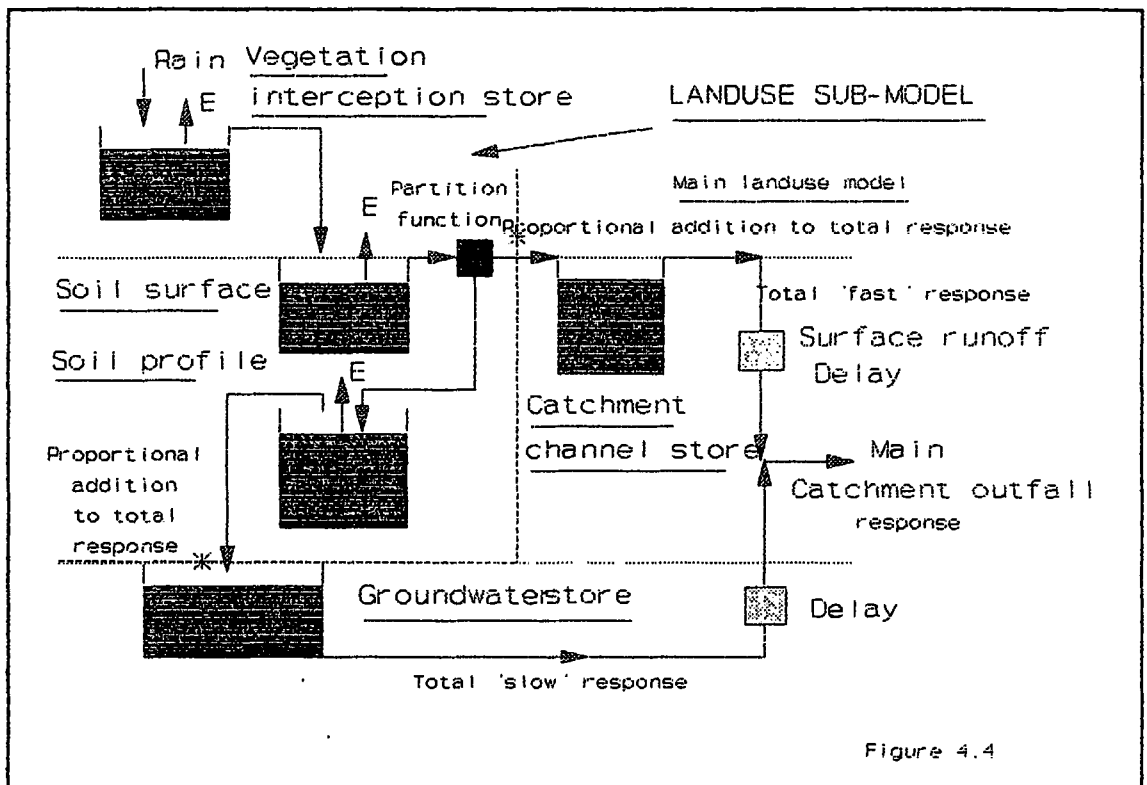


Figure 4.4

The module for each land use section has an interception or surface detention store, and a store representing the soil profile. The surface runoff from each land use is weighted by area and combined as an input to the channel store. The outputs from the channel store are rapid response runoff and are subject to a delay; these are then combined with the delayed groundwater store output to form the outflow, or runoff, from the catchment.

The effective rainfall is partitioned between surface runoff and infiltration to the different land use soil stores.

This infiltration combined with estimated transpiration losses controls the soil moisture deficits which develop under the different vegetation covers.

Percolation to groundwater from the soil stores occurs only when the soil is in a free draining condition, that is when field capacity has been exceeded as with equation 4.11. These percolates are weighted by land use area and combined as one input to the groundwater store. The output from this store, equation 4.15, which is considered as a non-linear reservoir, is given a delay and combined with the surface runoff for that time interval to form the total runoff from the catchment.

The system initial state parameters, p_{13} , p_{14} and p_{18} , are set at nominal values to begin the optimization and re-optimised at the end to 'fine tune' the total catchment volume output for the surface and groundwater components. This also improves the hydrograph fit between the dry initial start to the simulation and the wet winter conditions when the soil stores exceed field capacity.

Table 4.1

Parameter name	Description	Remarks
Interception store:-		
p_1	Size of interception store (mm). $0 < p_1 \leq 5$	Detention store for grasslands.
p_2	Penman evaporation factor for interception store. $1 < p_2 \leq 2$	Model designed for Penman data, but will use other evaporation data.
p_3	Interception store contents (mm). $0 \leq p_3 \leq 5$	System initial state.
Surface soil and detention store:-		
p_4	Size of surface detention and soil moisture store (mm). $0 < p_4 \leq 10$	Litter and humic layer.
p_5	Penman transpiration factor. $0.3 < p_5 \leq 1$	Penman E_0 or E_T .
p_6	Store contents (mm). $0 \leq p_6 \leq 10$	System initial state.
Surface runoff store:-		
p_7	Partitioning factor between surface runoff and infiltration. $0 < p_7 \leq 1$	Surface runoff estimated and residual is infiltration.
p_8	Exponential factor relating surface runoff to soil moisture deficit. $0 < p_8 \leq 1$	Reduction factor.
p_9	Exponential factor relating surface runoff to intensity of precipitation input. $0 < p_9 \leq 1$	Increase factor.
Channel routing store:-		
p_{10}	Channel routing store exponential. $1 < p_{10} < 3$	Determines shape of release curve.
p_{11}	Channel routing store factor. $0.1 < p_{11} \leq 1$	

P ₁₂	Runoff delay. $p_{12} \geq 0$	In model time intervals.
P ₁₃	Initial volume in transit (mm). $p_{13} > 0$	System initial state.
P ₁₄	Initial contents of channel routing store (mm). $p_{14} > 0$	System initial state.

Soil moisture store:-

P ₁₅	Transpiration ceases at this deficit (mm). $p_{16} < p_{15} < 200$	Wilting point.
P ₁₆	Transpiration starts to decrease at this deficit (mm). $0 \leq p_{16} < 50$	Moisture stress point.
P ₁₇	Percolation to groundwater factor. $0 < p_{17} \leq 1$	Percolation only occurs when soil profile is above 'fieldcapacity'.
P ₁₈	Initial soil moisture deficit (mm). $0 < p_{18} < 50$	System initial state.
P ₁₉	Maximum deficit below forest cover when transpiration ceases (mm). $80 \leq p_{19} < 200$	Optimised value removes store contents bias.
P ₂₀	Maximum deficit below heather when transpiration ceases (mm). $40 < p_{20} < 80$	- as above -

Groundwater store:-

P ₂₁	Denominator of groundwater store contents fraction. $30 < p_{21} < 300$	Equal to store contents, G_s , when changing at the rate of 1 mm per time interval.
P ₂₂	Groundwater store exponential. $1 \leq p_{22} < 4$	Sets curvature of recession curve.
P ₂₃	Groundwater delay. $p_{23} \geq 0$	Plus 1 data time period.

**Calder/Newson empirical parameters:-
(Process studies values in brackets).**

P ₂₄	Sets canopy saturation limit. $0.01 < p_{24} < 0.07$ (0.045)	Limits value for interception loss.
P ₂₅	Forest transpiration factor. $0.1 < p_{25} < 1$ (0.910)	

- P_{26} Forest interception factor.
 $1 < P_{26} < 10$ (6.990)
- P_{27} Forest interception exponential factor.
 $0.001 < P_{27} < 0.2$
 (0.099)
- P_{28} Heather transpiration factor.
 $0.1 < P_{28} < 1$ (0.500)
- P_{29} Heather interception factor.
 $1 < P_{29} < 5$ (2.650)
- P_{30} Heather interception exponential factor.
 $0.1 < P_{30} < 0.5$
 (0.360)

Catchment vegetation and open water area proportion factors:-

- $A_g \geq 0$ Proportion of grassland.
- $A_f \geq 0$ Proportion of coniferous forest.
- $A_d \geq 0$ Proportion of deciduous forest
- $A_h \geq 0$ Proportion of heather moorland.
- $A_w \geq 0$ Proportion of open water.

where $A_g + A_f + A_d + A_h + A_w = 1$

Chapter 5.

Model fitting techniques.

5.1 Systematic approach to parameter adjustment.

When the objective of the simulation has been defined and the appropriate model chosen then the next process is to fit and validate the model on observed flow data from the catchment. The method of fitting is dependent on the model chosen and on the data available. If the model is physically based then the quantification of, for example, surface runoff from precipitation input and catchment conditions will be based on field observations. In theory, at least, the parameters involved in the mathematical expressions should not require optimization. If, as is typical, the relationship used is a general one and not specifically derived from an experimental study in the field, or a conceptual abstraction representing an area, it will be necessary to iteratively adjust and test the parameter values. This is done until the model gives an acceptable approximation to the observed stream flow.

The term 'optimization algorithm' refers to the more sophisticated 'automatic' approach to the process of finding optimal values for the parameters rather than the simple 'trial and error' technique where the operator changes a value and studies the result on computer graphical or numerical output. This latter interactive method is still quite often found in the field of hydrological modelling. Its only advantage is that it allows the modeller complete control of the search for an optimum. However, it is not feasible to use this method for the complex highly parameterized model.

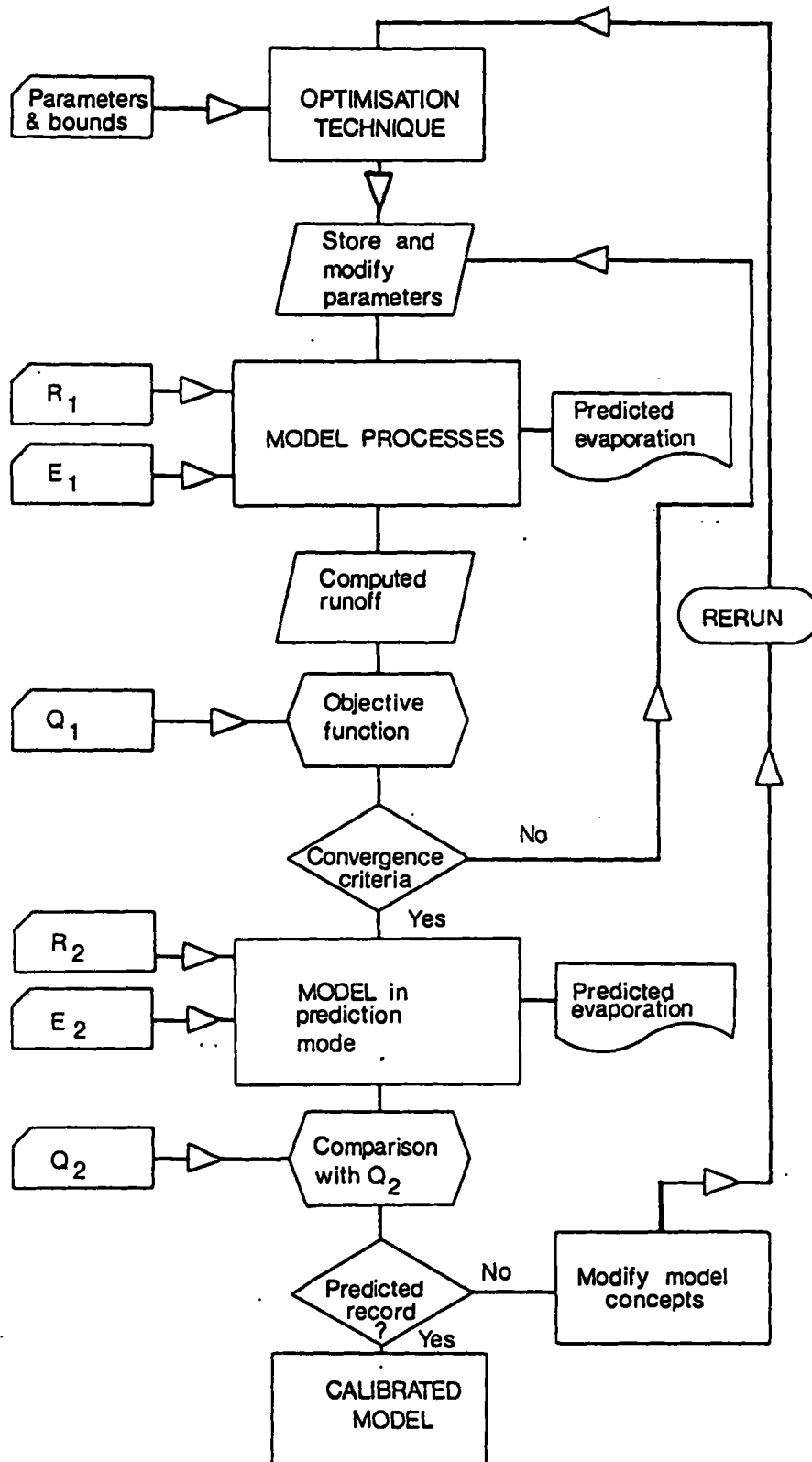


Fig. 5.1 General schema for split record calibration and validation of a model. The observed flow record is split into two periods, Q_1 and Q_2 , together with the rainfall inputs, R_1 and R_2 , and the evaporation 'loss' inputs, E_1 and E_2 .

The general outline of the fitting technique used in this case is shown in the system flow diagram shown in Figure 5.1 taken from Blackie and Eeles (1985). The optimization algorithm controls the process calling the model function as a subroutine; this was also shown by the process elements in Chapter 1 Table 1.1. The optimization routine first calls the model and using the input of estimated values of the parameters, together with their upper and lower bounds, the model is initialized to provide the starting point of the search for a minimum in n -space. This initial model is then applied to the calibration data, and the starting value of the objective function found. The optimization then passes back through the loop to the 'Store and modify' section of the diagram where this time the optimization technique modifies the initial parameters to begin the search from the initial parameter values. The process then passes through the optimization loop again with the model called as a subroutine. The calibration data, R_1 , E_1 and Q_1 are input during each call to the model either from direct access computer file storage or from computer random access memory (RAM) in the case of the HYRROM package. The latter is limited to five years of calibration data and fifty years of simulation data which uses a large amount of RAM. If the result of this displacement of parameter values by the optimization routine and subsequent model run through the calibration data set is a minimization of the objective function, then the new values are stored. If this is not the case and the objective function is larger, then the new values of the parameters are discarded and different values selected by the optimization search technique. This process continues until the criterion for convergence of the

model is met, and there is no significant reduction of the efficiency of fit, F_{relative} , calculated by the model from the objective function. Sometimes the search is terminated by the optimization algorithm when it cannot continue the search. If the criterion for convergence of the model is met then the model is regarded as partially calibrated and ready for validation on a longer set of data.

To validate the model it is then tested against a longer second period of observed data in simulation mode by assessing its performance with different sequences of input values from two new data sets of rainfall and evaporation, R_2 and E_2 , and its response to different sequences of changing store contents. The simulated flows are compared with the observed flows, Q_2 , and the success, or failure, of the model is gauged from the accuracy of the simulated flows. It is essential that the two periods are hydrologically homogeneous in response to rainfall and 'loss' function inputs, and have been tested for this condition beforehand by double mass analysis or a similar technique. Experience has shown that a fundamental change in the rainfall/runoff regime will invalidate and bias any modification to the hydrological model concepts that are made in response to defects in the simulation. The optimal values of the parameter set will try to give a mean fit to the data before and after the change if it occurs in the calibration period, and will be biased in relation to each simulation period. Gross and persistent errors of data and its processing may become apparent at this point, and it must be emphasized that the need for a very good data set which is compatible with the precision required of the model is of paramount importance in building a model which can be used to extrapolate the

results. Data processing errors should therefore not be ignored, or runs of data replaced using linear interpolation between the first and last data of the period.

If residuals are found from the observed data that stem from differences in timing or magnitude of flow events and these are unacceptable in terms of the accuracy required of the simulation by the model, then the concepts embodied in the model have to be reexamined and possibly modified, or discarded in favour of new concepts as shown by the return path in Figure 5.1. The model has then to be recalibrated and verified to test its modified form.

However, if there is a successful comparison then the model is regarded as fully calibrated and can be used to achieve whatever was initially set as the objective(s) of the simulation. Hydrological models are not 'reality' and great care has to be taken in assessing the limits of their use in a particular area or extrapolating the results to a different area or hydrological regime. This is particularly so if attempting to extrapolate a model calibrated on one area to another area which is ungauged.

5.2 The split record and objective functions.

Regardless of the numerical algorithm used for fitting the model, the systematic approach outlined in section 5.1 has become widely accepted for hydrological modelling using large data sets. This method where the observed record is long enough to split into a shorter calibration period and a longer period for validation has become known as the 'split record' test. It is valuable in the analysis of the simulation results because the second period provides different sequences of input

data, which form 'loss' from or input to the model stores with different store contents and producing values of flow output that are not unique to a particular sequence input data. This flexibility is essential in the field of hydrological modelling and has led to the very limited success of linear models which imply a direct input/output response, Dooge (1973).

In both fitting and validating the results of a simulation some objective criterion is needed to compare the observed and simulated stream flows. The fitting of parameters is usually based on the minimization of the objective function, F , the sum of squares of the residuals between the simulated and observed values as given by:-

$$F = \sum_{i=1}^n [(Q_s)_i - (Q_o)_i]^2 \quad (5.1)$$

where Q_s are the simulated flows and Q_o are the observed flows for the time interval i taken over all the n ordinates. This same objective function is then used to validate the model using a set of data from a different period.

This expression does tend to place an undue weight on peak flows as mentioned by Douglas (1974), and reduce the effect of low flows of less than 1 mm in a **short** run of data. Fitting a curve by least squares has the effect of smoothing out peaks and troughs. However, the latter make up such a large percentage of the total volume of flows in **long** runs of data, as shown in the data analyses in Chapter 8 that the few large flow events lose their dominant effect. The sensitivity of the rapid response parameters affecting the surface flows can be 'damped' by this mass of low flow data. In setting up the

basic structure of the parameter set between 'rapid' (surface response) and 'slow' (sub-surface) response parameters the author has found it useful to use a much higher even power for the optimization of the rapid response parameters:-

$$F = \sum_{i=1}^n [(Q_o)_i - (Q_p)_i]^8 \quad (5.2)$$

This has the effect of concentrating the optimization on the rapid changes in flow brought about by the surface runoff parameters, and severely curtailing the effect of the mass of groundwater low flow data. Previously it has been the practice to select a short sub-set of data where the rapid response flows are dominant but in Eeles and Blackie (1993) better results were obtained using the full run of calibration data with the eighth power objective function.

It was thought at an early stage in this research that a proportional objective function would take out any stress on the magnitude of the high stream flow events:-

$$F = \sum_{i=1}^n \frac{[(Q_o)_i - (Q_p)_i]^2}{(Q_o)_i^2} \quad (5.3)$$

However, this normalization was found to increase the dominance of the low flow events to such an extent that generally for data sets high flow events had no significant effect on the objective function. This led to the conclusion that function 5.3 would be most useful in optimizing the slow response parameters affecting the baseflow coming from the groundwater store in periods when the streamflow recession was dominant.

A simple logic switch is used to select the objective function for the whole calibration period.

In the early modelling work only small data sets were used due to limited computer resources. However, it is now the practice to use data sets from eighteen months to four years to calibrate a model. This highlights the disproportion between the high total volume of low flows and flows from storm events, and makes optimization of all parameters difficult if the least squares objective function is used. However, this function is still very useful in assessing the effects of optimization on the model explained variance and correlation when comparing the calibration and validation periods.

In the nine parameter model, HYRRROM, only the least squares function is used. However, the criterion for convergence in section 5.9, based on the relative fit of the model, is taken to assess whether the changes in F are significant. Eeles and Blackie (1993), for the more complicated land use model, needed the focused objective functions offered by the use of the higher even power for peak flows, and the proportional version used for low flows.

5.3 Sensitivity of parameters and response surfaces.

In the work by Johnston and Pilgrim (1973) and later by Pickup (1979) an attempt was made to assess the relative sensitivity of model parameters by displacing their values by $\pm 50\%$. As can be seen later in the discussion of response surfaces this approach can be disastrous if local optima are found and large displacements made from that point. To establish the sensitivity of a parameter in the objective function space only a small variation should be made, and then

the sensitivities found are only really relevant to the initial values of the parameters at the given point. This $\pm 50\%$ method for each optimized parameter in F-space established by the parameters selects an initial value about which the large variation may be meaningless in terms of the general n-dimensional topography. Sensitivity can be defined as the magnification of a change in the simulation caused by a change in the parameter values or data, but since the model functions are mainly non-linear the sensitivity varies with time. Two examples of data processing error leading to displacement of the model simulation from the observed hydrograph are given by Eeles (1978). One was caused by a non-linear error in processing rainfall data, and another by error using a wrong planimeter setting while measuring areas from net radiation data.

In the early use of a similar model to HYRRROM a recommended initial set of four parameters to optimize was found by the method of $\pm 50\%$ displacement to obtain the 'most' sensitive parameters. Sensitivity in this case was defined by the relative ranking of the parameter in decreasing the objective function. In order of sensitivity these are:-

1. P_5 the Penman transpiration factor.
2. P_{22} the power in the groundwater release function.
3. P_7 the surface runoff/infiltration partitioning factor.
4. P_1 the depth of the vegetation interception store.

A different set of parameters, in which 1 and 2 retained their position and 3 and 4 were exchanged, was obtained using the same displacement method and HYRRROM but different data from Zambia, Son Il (1991). Using a $\pm 5\%$

difference and four series of three year Kenyan catchment data with the different Blackie nine parameter model, Blackie (1972) obtained a totally different rank for the same parameters: P_5 came fourth, P_{22} was still second, P_7 came third and P_1 was last.

Taking the general model equation that Q_s depends on the model function of p_{opt} , R and E :-

$$Q_s = f(p_{opt}, R, E) \quad (5.4)$$

where Q_s is the simulated flow, p_{opt} is the set of optimized parameters, and R and E are the sets of rainfall and evaporation data.

From 5.2 and considering the values of the objective function for particular points, it would seem that the sensitivity in terms of simulation model sub-functions, f , could be directly defined with different values to the set of parameters as:-

$$S_{p_i} = \frac{f(p_i, R, E)_j - f(p_i + \delta p_i, R, E)_{j+1}}{p_i - (p_i + \delta p_i)} \quad (5.5)$$

$$= \frac{\Delta f_j}{\Delta p_i}$$

for the i -th parameter of the set of n parameters

where j is the model function f corresponding to the i -th. parameter, and $j+1$ is the value of f for the displacement of the parameter. S_{p_i} is the sensitivity of the parameter p_i for the displacement δp_i and is shown as the ratio of the difference between the change in the model function f for the small change in p_i . However, this is unsatisfactory generally as the value of the sensitivity will vary as the value of p_i .

At the optimal value the sensitivity would be zero or hardly change while with higher values of f it would change with every variation of the topography.

With the Taylor expansion of degree 1

$$f(p_1 + \delta p_1, R, E) \approx f(p_1, R, E) + \delta p_1 \frac{\partial f}{\partial p_1} \quad (5.6)$$

$$f(p_1 + \delta p_1, R, E) - f(p_1, R, E) \approx \delta p_1 \frac{\partial f}{\partial p_1}$$

However, this is still dependent on the input data set.

A three dimensional response surface is shown in Figure 5.2, taken from Blackie and Eeles (1985), for 'RDEL' the rapid response component delay, p_{12} , and 'GDEL' the slow response component delay, p_{23} in hours. This illustrates the problems in taking any measure of the sensitivity of a parameter in feasible space with precisely defined boundaries. It would appear that by limiting p_{12} to one to nine hours, based on field observations of stream rise after a storm, the global minimum has been missed as it probably lies outside the feasible space. The 'peak' at 65 to 85 hours on the p_{23} axis shows that the limit for this parameter was set too high at 175 hours. The ripples along the topography parallel to this axis reflect the cyclical variation of the input data. An interesting use is made by Kuczera, 1990, of response surface plots to assess hydrologic model nonlinearity by comparison of the elliptical contours about the optimum with linearized probability regions. Kuczera also draws the important conclusion that the degree of nonlinearity is not solely an intrinsic property of model structure; it also depends on the information available to

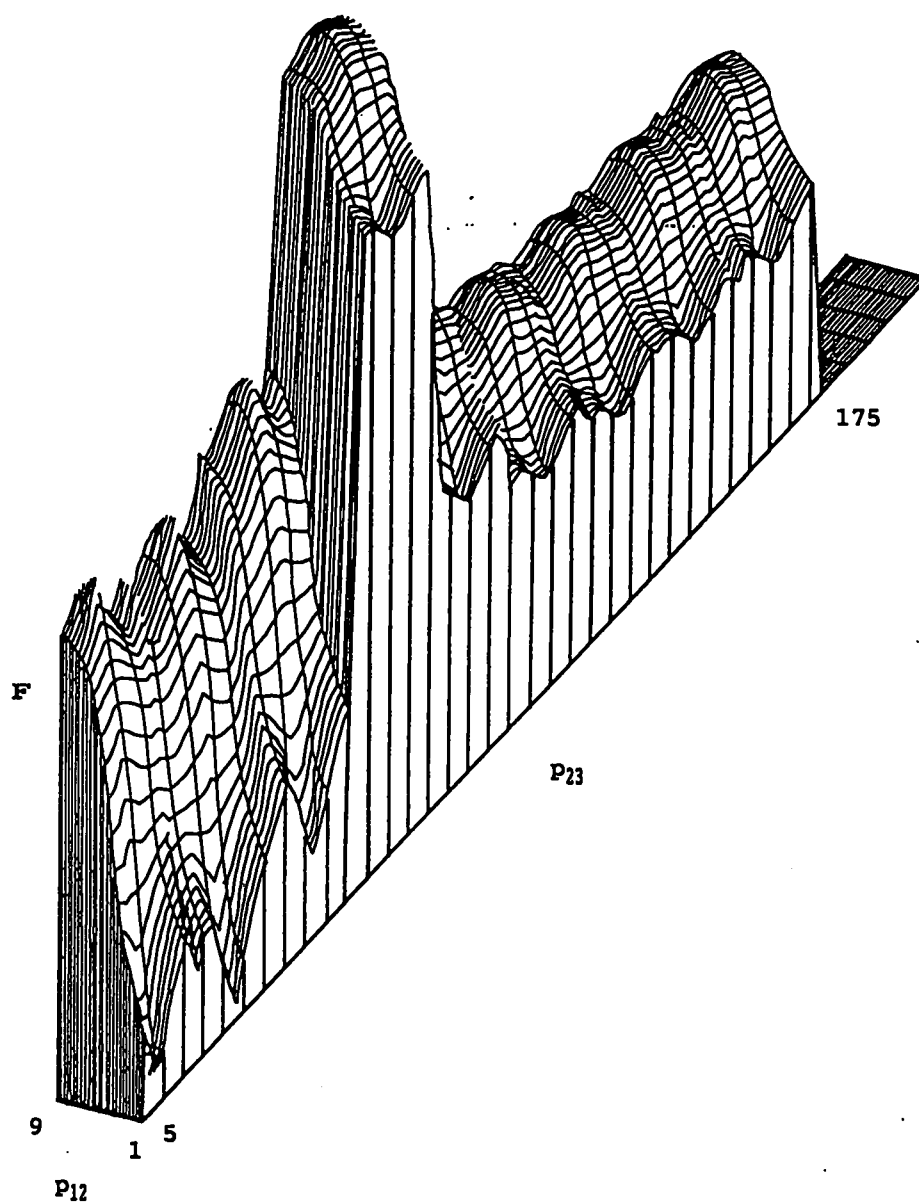


Fig. 5.2 A three dimensional mapping of F-space for the routing parameters p_{12} (RDEL) and p_{23} (GDEL): the surface runoff delay and groundwater delay. The parameters are subject to constraints: $1 \leq p_{12} \leq 9$ hours and $5 \leq p_{23} \leq 175$ hours. The base grid dimensions represent time intervals of 20 minutes for p_{12} and 5 hours for p_{23} .

infer model parameters. This would imply that consideration of the parameter sensitivity will vary between data sets.

A statistical discussion is given by Jones, 1983, of 'empirical' models fitted by optimization. Jones rejects the construction of an explicit statistical model as a difficult, if not impossible, solution to the analysis of optimization results. The expressions developed are similar to the basis of the maximum likelihood theory and its discussion by Cox and Hinkley, 1974, and relies on similar assumptions: at least the first three derivatives of the model function exist and the statistical parameters between data sets are the same. Unfortunately, the latter assumption is not likely to hold if both data sets used for calibration and validation of the model are large sets running over several years. Thresholds for particular parameters invalidate these assumptions and the statistical theory is not generally applicable to some conceptual models.

5.4 Strategy of parameter optimization.

In general the type of conceptual model most successful in simulating the rainfall/runoff response of a complete catchment has between eight and twenty parameters: in the case of HYRRROM there are nine parameters which can be made active but there are ten more which are set at fixed values. Since the parameters cannot be said to be completely independent, due to the model input/output structure of each module, it has been found empirically that the optimization of sub-sets allows control of the parametric structure to be maintained. Making all parameters active at the same time often leads to a very distorted structure in which an initial parameter optimized

towards either of its limits distorts the 'lower' parameters reducing their sensitivity by the reduced 'effective' rainfall/runoff information affecting the objective function.

For both the models described in Chapter 4 the parameter set is first initialized with values which are physically realistic or derived from field observations. These are then divided into three sub-sets so that each conforms to an heuristic optimization strategy in which the experience gained during the floating of one set provides the basis for the subsequent work. This approach divides the parameters into sets controlling the long-term volume response from groundwater storage, short-term surface runoff and the contents of stores.

The long-term approach usually considers the model output at intervals of a month, with the observed daily data totalled for the month, and the objective function applied on a monthly basis.

The parameters involved in this optimization on monthly totals are the 'loss' factors, p_2 , p_5 and p_{17} , the partitioning factor, p_7 , and the groundwater release parameters, p_{21} and p_{22} , as described in Table 4.1. The parameters involved with the rapid surface response and soil moisture percolation form the second set and are optimized at the data frequency. These parameters are for surface runoff, p_8 and p_9 , the channel store release parameters, p_{10} and p_{11} , unrestricted transpiration p_{16} , percolation p_{17} , together with the two delay factors p_{12} and p_{23} . These parameters govern the shape and fine structure of the simulated hydrograph whilst the rest affect the area under the graph and are also optimized at the data frequency. The importance of the storage parameters has to be kept in mind as a store can empty under the action of the model processes and

then only fill with difficulty. Alternatively, the soil and groundwater stores can continue to fill, and as there is no limit on their size can act as 'sinks' affecting the total volume of flow over the calibration period. The size of the interception store, p_1 and soil surface store, p_4 , also provide time interrupt controls while they fill and so delay the surface runoff response. The initial state parameters can be very important as controls of the initial model prediction. When a good total volume agreement has been achieved, the initial volume bias may make up the bulk of the volume error at the end of a simulation. In selecting the order of the parameters made active in each set the sequence of the mathematical functions within the model algorithm is usually the deciding factor, but this can be overridden by the perceived sensitivity of a parameter.

Unfortunately, if the volume approach is used two sets of parameters can be obtained which refer to the monthly and daily data sets. The models are incompatible and no further progress can be made with optimization of these two parameter sets. This case is fortunately rare, and experience has shown that the only solution is to choose the most physically realistic values from each set to form a third set. The third model parameter set has then been found to make an improvement in the simulation fit.

After several iterations of this strategy and when a satisfactory simulation has been achieved, then, to refine the detail of the simulation, a final attempt is made at optimization of a parameter set whose elements consist of the most important, or sensitive, parameters from each module as described in section 5.3. However, this 'fine tuning' of the

model parameters can fail due to the poor quality or lack of information contained in the data. The goodness of fit cannot be better than the inherent errors and bias of the data set.

5.5 Problems arising from the model structure.

The model algorithm with its structured sequence of conceptual functions is divided into modules representing the sequential operation of processes in the hydrological cycle. The reduction of the initial precipitation input by the sequence of inputs to the stores and the evaporation loss functions and inputs to routing functions has the result that there is less rainfall/runoff information available to successive functions. This is shown in the reduction of the effectiveness of parameters as they control and operate in top/down sequence on these reduced outputs. The decreased information available from the observed data is not sufficient to infer the parameter optimal value with any degree of accuracy. Changes in these parameters may not therefore significantly affect the objective function in the calibration period.

An example of this effect can sometimes occur with the two paths in the algorithm: the rapid response of the surface runoff and slower infiltration into the soil. Optimization of the percolation to ground water parameter, p_{17} , in the same iteration as parameters in the surface runoff and channel routing stores can lead to the effective removal from the model of the groundwater module by greatly reducing the flow through the soil store. This is one reason for the fixing of this parameter in HYRRROM as well as the reduction of the number of active parameters.

The parametric structure of the models varies greatly between the nine and thirty parameter models with some parameters being common to both sets of modules from the routing and storage concepts. This is shown in Figs. 4.2 and 4.3 of the preceding chapter.

The parametric structure of the two models also varies if the time interval of the data available is only monthly. This is a very common case with projects and data from abroad. The monthly data has then to be distributed on a daily basis for processing by the model algorithm which leads to gross errors when the simulated monthly hydrograph is changing rapidly. However, the overall volume changes for the period are reasonably accurately simulated, and in the periods when groundwater flow is dominant give a good monthly representation.

5.6 Feasible function space.

The function-space spanned by the parameters has as many dimensions as there are parameters provided that all the parameters are sufficiently independent to prevent collapse to a lower dimensionality. Thus to restrict the search to probable regions and maintain the physical relevance of the parameters the search of this n-dimensional hyperspace has to be restricted by the setting of bounds. These are physically based constraints derived from theory or field observations relevant to the model concepts. In HYRRROM the boundary limits are arbitrarily set at $\pm 50\%$ of the initial value of the parameter chosen. This setting of constraints effectively limits the number of optimization iterations, and model

processing runs, by restricting the area of search as shown by the sharp cutoff 'cliffs' in Figure 5.2.

However, the model is not the 'real' world, but only an abstraction from it, and the bounds may have to be relaxed as they are somewhat arbitrary in terms of restricted model concepts. This relaxation of the search limits may allow a particular response surface contour to be crossed which is a barrier to the optimization search vector. From experience when this has been allowed the search has continued outside the boundary but the search direction has turned and the optimal parameter has then been produced in the original feasible space.

5.7 Constraints, scaling and criteria for convergence.

This problem of constraints defining the boundaries of feasible space is a general one with any optimization technique. An obvious solution to this is to allow the search to continue up to the boundary and then halt the search at this point with a reversal of the search vector so that the next step is away from the boundary. This leads to the possibility of continual returns to this boundary point, but a limit can be applied to the number of times the search may approach the boundary in one iteration. The use of penalties or penalty functions was considered but rejected due to the possibility of creating a false optimum at, or near, the boundary.

Another problem is the weighting given by size of a parameter. The range that parameters can have is shown in Table 4.1 where some parameters have zero as a lower bound and others can have values as great as 300. In Pickup, 1979, the parameters were simply scaled between 1 and 10 which still

allows for large relative differences at either end of the scale.

A solution to these parameter constraint problems for the two models was found by using a scaling function based on the upper and lower bounds. This $\sin^2\theta$ function scaled the parameters and transformed the problem from a constrained to an unconstrained one as follows:-

$$p_{scaled} = \frac{p_i - p_{lower}}{p_{upper} - p_{lower}} = \sin^2\theta_i \quad (5.7)$$

this ensures that $p_{scaled} \in [0,1]$ provided that $p_i \in [p_{lower}, p_{upper}]$

$$\theta = \sin^{-1} \sqrt{\frac{p_i - p_{lower}}{p_{upper} - p_{lower}}} \quad (5.8)$$

p_i is the parameter being allowed to 'float', p_{lower} is its lower bound, and p_{upper} is the upper bound.

The advantage of this transformation is that for any value of θ_i then $p_i \in [p_{lower}, p_{upper}]$. Within these bounds the parameters are scaled between 0 and 1 by the $\sin^2\theta$ function.

The criterion for convergence of the optimization had originally been that when all parameters were changing by less than a given percentage using a FORTRAN subroutine AMAX. However, the author found from experience that different sets of parameters could be optimized from different sets of initial values, and these gave the same approximate degree of optimum fit to the observed hydrograph. So the criterion for convergence was tried as when equation 1.3 of Chapter 1 was changing by less than a set amount, δ :-

$$(F_{relative})_{j+1} - (F_{relative})_j < \delta \quad (5.9)$$

where $F_{relative}$ is the j -th explained variance,

$$(F_{relative})_j = \frac{F_{var} - F_j}{F_{var}}$$

and F_{var} the data variance with F_j the j -th reduced objective function with $F_j > F_{j+1}$

$$(F_{relative})_{j+1} = \frac{F_{var} - F_{j+1}}{F_{var}}$$

Combining the three equations the expression for δ becomes

$$\frac{F_j - F_{j+1}}{F_{var}} < \delta \quad (5.10)$$

where $0 \leq \delta \leq 0.01$, F_{var} is the data variance, and F_j and F_{j+1} are the objective functions for two successful consecutive iterations: $F_j \geq F_{j+1}$ and $(F_{relative})_{j+1} \geq (F_{relative})_j$. This function is used for testing convergence in both models, but δ is set from experience to 0.03 in the HYRROM algorithm.

5.8 Mathematical formulation of the search technique.

The search for an optimum, F_* , is made in the least squares objective function space:-

$$F_* = \sum_{k=1}^n [(Q_o)_k - (Q_*)_k]^2 \quad (5.11)$$

[generally $F_* \neq 0$]

and stop when

$$\frac{F_j - F_{j+1}}{F_{var}} < \delta \quad (5.12)$$

The parameters, p_i , span the n-dimensional feasible space of the objective function, but F_* is not necessarily the global minimum value but only satisfies the criterion for convergence given by equations 5.11 and 5.12. This simply expresses the fact that the fit to the observed hydrograph is within previously chosen acceptable limits, and the model parameters obtained by this calibration have to be confirmed by a pass through the validation period of data to give:-

$$\varepsilon = |(F_{relative})_* - (F_{relative})_{validate}| \quad (5.13)$$

The acceptance or rejection of ε becomes subjective; ideally ε should be zero but this is too much to expect from a longer run of data with all its inherent errors. The simulated hydrograph has to be examined in relation to the observed one to see where there are large divergences. These can be caused by deficiencies in the model algorithm or concepts, observer error, instrument error, or data processing error. If there are large consistent divergences then the model concepts may require modification.

Chapter 6.

Optimization algorithms available in the field of Hydrology.

6.1 Availability of algorithms.

To say that the algorithms covered in this chapter are 'in use' is to imply wide application throughout the hydrologic modelling field. It would be correct to say that they are available but not used very often, and when used they are rarely applied with knowledge of the method and its limitations. A particular algorithm is applied simply because it happens to be easily available, and not because it is the most suitable for the particular problem.

At the beginning of this research the only algorithm available, and in common use, at the Institute of Hydrology was a very modified version of the Rosenbrock rotation of axes algorithm. This was slow in operation with interminable searches of the objective function space which produced insignificant changes of parameters, but little or no improvement to the simulated hydrograph. The author later introduced a version of the Nelder and Mead Simplex algorithm, as well as greatly improving the efficiency of the Rosenbrock algorithm and data format for input/output with an overlay of the data month by month. The work was soon transferred from the mainframe to an IBM AT personal computer, which made it imperative to organise the algorithms and their program coded efficiently to mount them on the small computer.

All the algorithms tested in this research had defects in their FORTRAN coding. The only code mounted, and running, on the Institute mainframe computer was the Rosenbrock algorithm, and all published code used in this research needed

modification and correction by the author of this thesis before it would even run. This work had to be undertaken as well as interfacing the algorithms to the models and data before any comparative tests could be made for research.

The comparative tests made use real data from observations taken in experimental catchments. A detailed analysis of the reliability and homogeneity of the data, and comparison between catchments, is made later in Chapter 8 because of its importance to the extrapolation of the results of this research. The results are empirical and strictly apply only to the data used, but since the data is taken from a wide variety of sources the conclusions should at least provide a guide to the future use of the algorithms.

6.2 The concept of the direct search algorithm.

The direct search technique is one in which the strategy is to determine a search direction, minimize, or at least reduce, the function in this direction and then, having moved to this point, to choose a new search direction. This process is repeated until convergence is achieved. Any direct method, if it is successful, will yield the same optimal solution. The main variation is in the work done to obtain the optimal solution. The step lengths in the direction of the search, and directions of search are fixed heuristically, or by a particular strategy, rather than following an optimal schema. The number of search trials tends to be greater than with a slope dependent algorithm such as the Marquardt version tested in Chapter 9. This produced its optima in a remarkably small number of trials.

The alternating variable direct search algorithm used in this research is described in the following paragraphs. It was developed by the author from the very successful strategy employed in the first iteration of the Institute of Hydrology version of the Rosenbrock algorithm.

The alternating direct search method is the simplest in that it seeks the optimum value of each parameter in turn parallel to the axis of that parameter with the other parameters held at constant values. It is therefore relatively easy to program with minimal use of computing resources, but tends to converge very slowly. The main assumption is that each search direction will reduce the value of the objective function until the optimal solution is reached. The objective function space is built up from:-

$$F_{i,j} = \sum_{k=1}^{k=m} (Q_0 - Q_i)_k^2 \quad (6.1)$$

where $F_{i,j}$ is the value of the objective function for the j -th change, δ_j , in the parameter, p_i , which is allowed to float to give the simulated value of flow Q_i ; Q_0 is the observed value of stream flow for the k 'th time interval, and m is the total number of intervals.

The perturbed parametric vector, p , with n elements is:-

$$p = [p_1, p_2, \dots, p_i + \delta_j, \dots, p_n]^T \quad (6.2)$$

where T denotes the transpose of the vector.

The simplest form of search is to vary p_i and look for

$$F_{i,j} \geq F_{i,j+1} \quad (6.3)$$

for each parameter p_i in turn $i = 1, \dots, n$

For one or two parameters this search can be done on an interactive trial and error scheme but it is faster and more accurate if done on a semi-automatic basis by digital computer. The efficiency of an optimization algorithm then depends on how quickly the move is made from the initial parameter vector, p_{init} , to the optimal parameter vector, p^* . If p_{init} is badly chosen then an efficient algorithm is needed to achieve the optimum solution quickly. The size of the step taken in each search direction may increase rapidly until the area of F-space where a local optimum lies is located such that $F_{i,j+1}$ is less than $F_{i,j}$. This area is then searched with smaller and smaller steps until the optimal point is localized to within the chosen criterion for terminating the search. Box, Davies and Swann (1969) proposed such a search scheme for the alternating variables method. Their recursion formula in terms of the parameter p_i is

$$\begin{aligned} p_{i,j+1} &= p_{i,j} + \Delta_j \\ &\text{while } F_{i,j+1} \leq F_{i,j} \\ &\text{where } \Delta_j = 2^j \Delta_{i,1} \end{aligned} \quad (6.4)$$

$\Delta_{i,1}$ is the initial step size in the direction of p_i such that F is decreasing. As long as the inequality holds this formula is applied, but when this does not hold at the k -th iteration the step size is reduced as

$$P_{i,j+1} = P_{i,k} - 0.5^{j-k} \cdot \Delta_{i,k} \quad (6.5)$$

$$\text{while } F_{i,j+1} < F_{i,j}$$

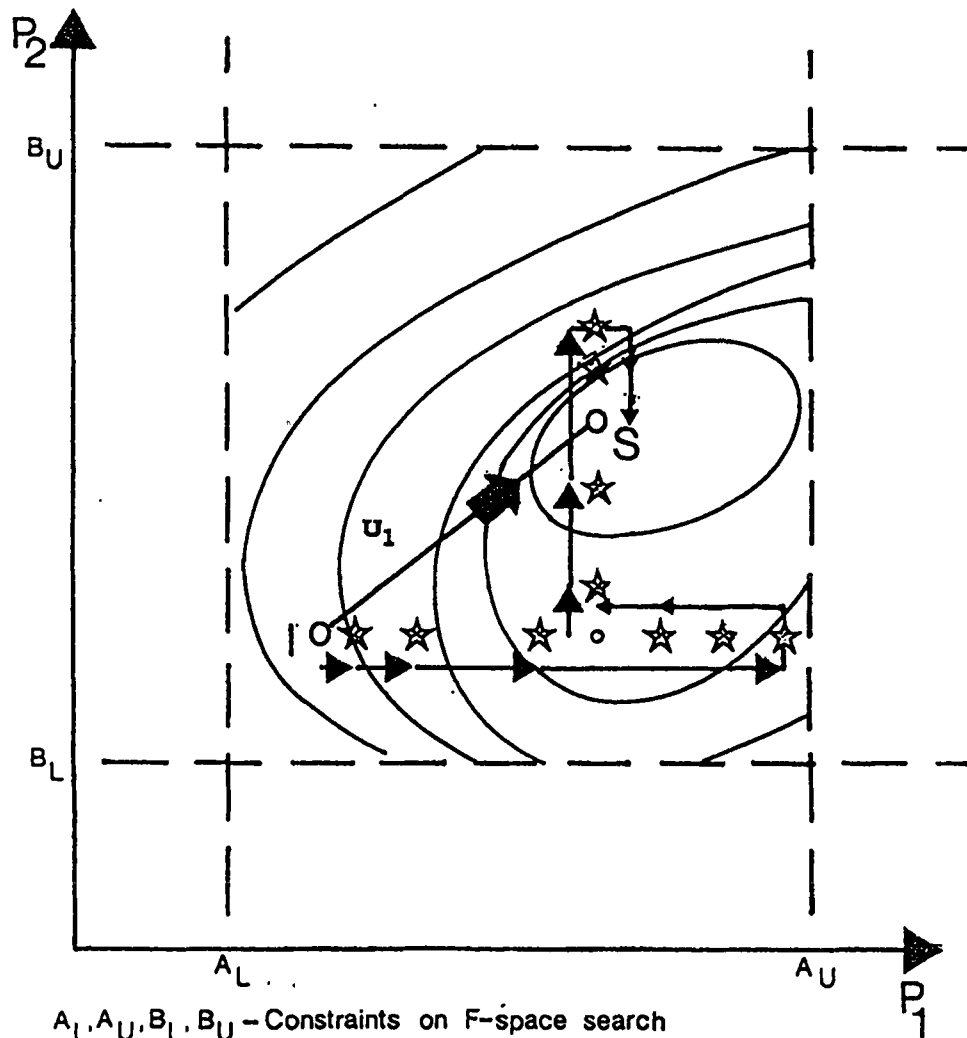
This search is continued as long as the inequality holds with the condition that $F_{i,j+1}$ is less than or equal to $F_{i,k}$ when the search terminates. This process is illustrated by Figure 6.1 which is a diagram of the Rosenbrock algorithm before the first rotation of axes.

6.3 The Gram-Schmidt orthogonalization of axes.

In Figure 6.1 the search made with the Rosenbrock algorithm is shown for an optimum in a two dimensional objective function space formed by two parameters. Rosenbrock (1960) used a similar scheme to the 'success/failure' concept above as the first stage of an optimization algorithm and then added a rotation of axes to continue the search. This rotation is made using the Gram-Schmidt orthogonalization process a discussion of which is given by Johnson and Riess (1982).

The code for this sub-routine was given by Ibbitt (1970) apparently based on the equations given in the Rosenbrock paper. This differed significantly in form and structure from the FORTRAN code for the same routine given in the version of Rosenbrock available at the Institute, but the basic equations, Johnson et al (1989) were the same. We compute an **orthogonal** set of vectors $[(U_i)^T \cdot U_j = 0 \text{ for } i \neq j]$ U_1, \dots, U_p where $U_1 = W_1$

$$U_i = W_i - \sum_{k=1}^{i-1} \frac{U_k^T \cdot W_i}{U_k^T \cdot U_k} \cdot U_k \quad 2 \leq i \leq p \quad (6.6)$$



A_L, A_U, B_L, B_U - Constraints on F-space search

☆ - Trial points in F-space

→ - Extending 'success' search vector

→ - Reversed search vector after 'failure'

○ - Final 'success' point for P_1 parameter axial search

IO - Search starting point and new origin for rotated axes

SO - Final 'success' point for first iteration of algorithm

→ - Resultant vector \underline{IS} aligned with new P_1 axis after translation of origin to point I and rotation of axes

Fig. 6.1

where W is a p -dimensional subspace of the vector space R^n and $[W_1, W_2, \dots, W_p]$ is any basis for W . In the case of the Rosenbrock algorithm the elements of W are the elements of $P_{iteration}$ the parameter vector found at the end of each iteration.

However, in view of the different handling of arrays and workspace in the two versions of the sub-routines and the warning in Johnson et al (1989) about the difficulties of expressing the process in computer code, the two sub-routines were tested by the author using dummy data and found to give similar results. The version which was then applied at the Institute being more concise in its handling of array elements and using less computer storage space was therefore chosen and continued to be used. No attempt is made when using this version to determine the most successful directions of search: the system of parameter axes is rotated to align the axis of the first parameter with the direction of the successful search vector for the iteration. All parameter axes are searched in the order in which the parameters are first 'floated' for optimization.

6.4 The Rosenbrock rotation of axes algorithm.

The original version of this search process was first put forward by Rosenbrock (1960). It was later identified by Ibbett (1970) as the most efficient algorithm of those tested for Hydrologic modelling, and has remained in general use in this field since then: particularly at the Institute of Hydrology with the version outlined by O'Connell et al (1970) and later by Clark (1973).

The version available at the Institute of Hydrology is one which uses the direct search algorithm given in section 6.1 for

the first iteration of searches parallel to each axis, and then searches parallel to each rotated axis for the minimum for that iteration. The method uses the 'success, success, failure, success' criterion for the termination of the search along any particular axis. The method is similar to the 'line search' one described by Box, Davies and Swann (1969), and the criterion for the above termination of the search parallel to any rotated axis is retained for rotation iterations after the first Rosenbrock rotation of axes.

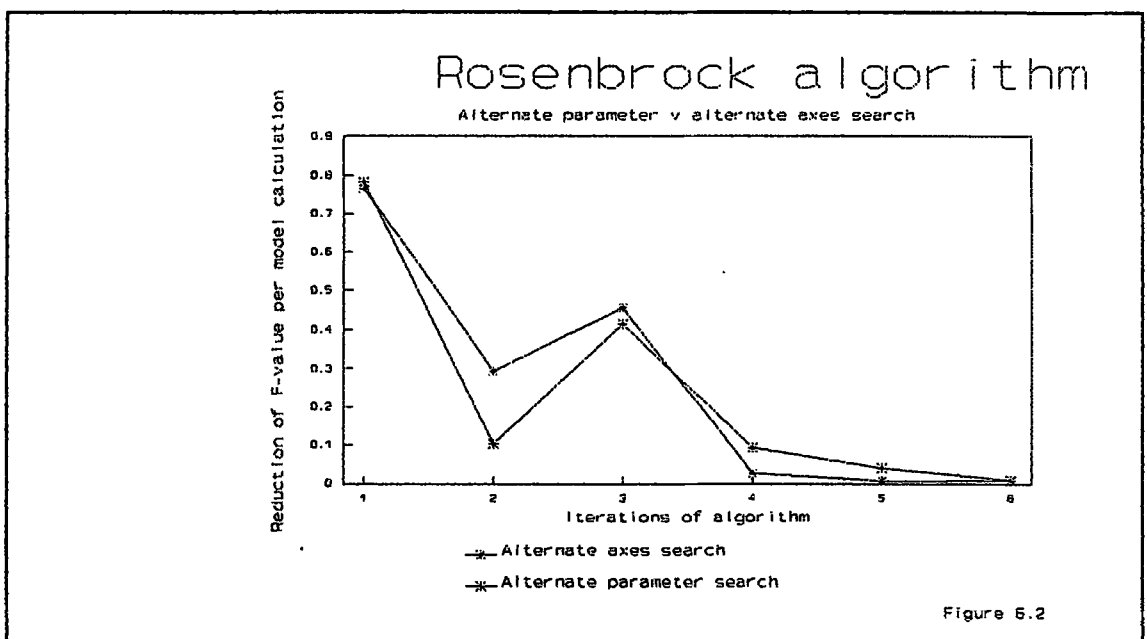
The axes are searched in turn by their initial order of parameter input, and the criterion for directing and terminating the search along a particular axis is the same as that which formed the basis for the direct search algorithm in Section 6.1. The strategy of directing the search vector is 'success', 'success', . . . , 'failure', . . . , 'success'. Here 'success' indicates a lower value of the objective function, and 'failure' a higher value than the last value. At the latter point the direction of search is reversed and the length of the search vector reduced by a factor of 0.5; this process continues until a value equal to or lower than the last 'success' value is reached when the search along that parameter axis is terminated. With descending monotonic values of the objective function given by repeated 'success' the step length of the search vector is increased by a factor of two. Rosenbrock originally suggested three as the factor but the even more rapid build-up in length can cause complications with any trigonometrical transformation of parameters: the parameter value becoming either negative, zero or exceeding the feasible upper or lower bounds.

At the end of the first iteration, when all the axes have been searched for the lowest value of the respective parameters, each complete search $i = 1, \dots, n$ gives a new U_k , then the origin is translated to the starting point and the system of axes rotated so that the axis of the first parameter is aligned with the resultant vector of the search in this iteration, as shown in Figure 6.1. The original version of Rosenbrock determined the most rapidly changing parameter in each iteration and aligned that axis with the resultant vector. It is difficult to visualize the rotation in n -dimensions, but it certainly requires less computing resources to simply rotate the system of parameter axes based on the first parameter each time as is done with the Institute version. It might be argued that this places undue importance on the first parameter, but since each axis is searched in turn the resultant vector at the end of the next rotation iteration should be the same by either method.

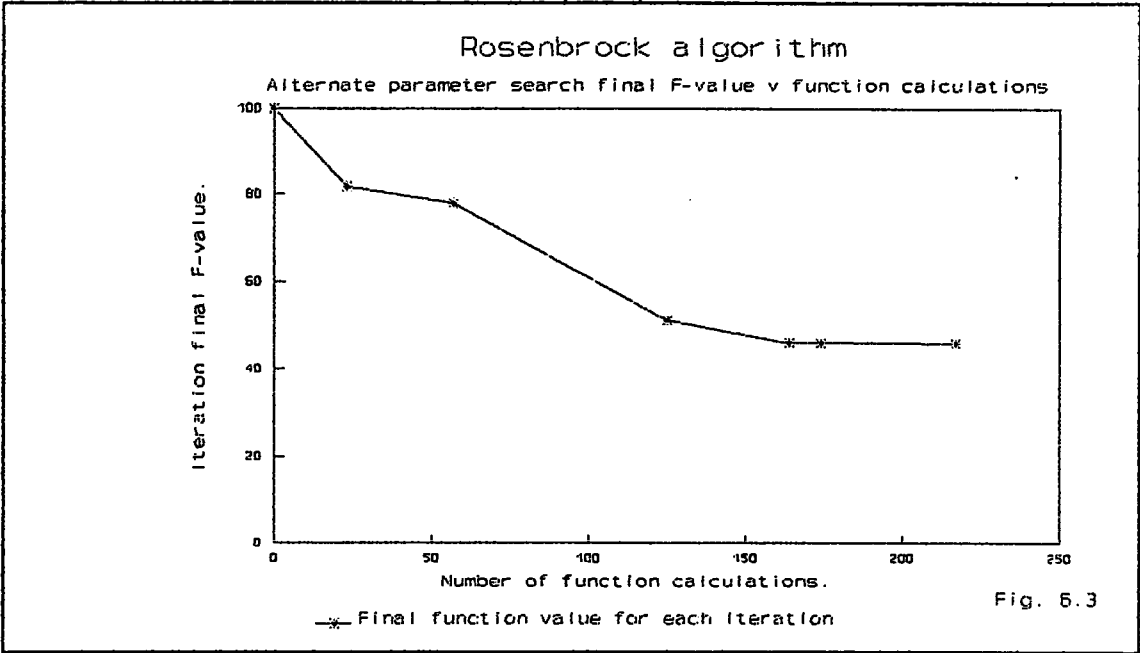
6.5 Step-wise application of the Rosenbrock algorithm.

Rosenbrock (1960) mentions the possibility that an alternative procedure to the one adopted by him of making a trial in each of the n directions in turn would be to make the number of trials depend on their success instead having n trials. Since this alternative was the procedure adopted in the current version of the algorithm used by the Institute of Hydrology it seemed necessary to test if there was any difference between the two Rosenbrock methods as Rosenbrock had made no systematic test of whether his suggested alternative procedure would affect the time taken to reach an optimum. The code for the alternate axes algorithm as set out in Section 6.3

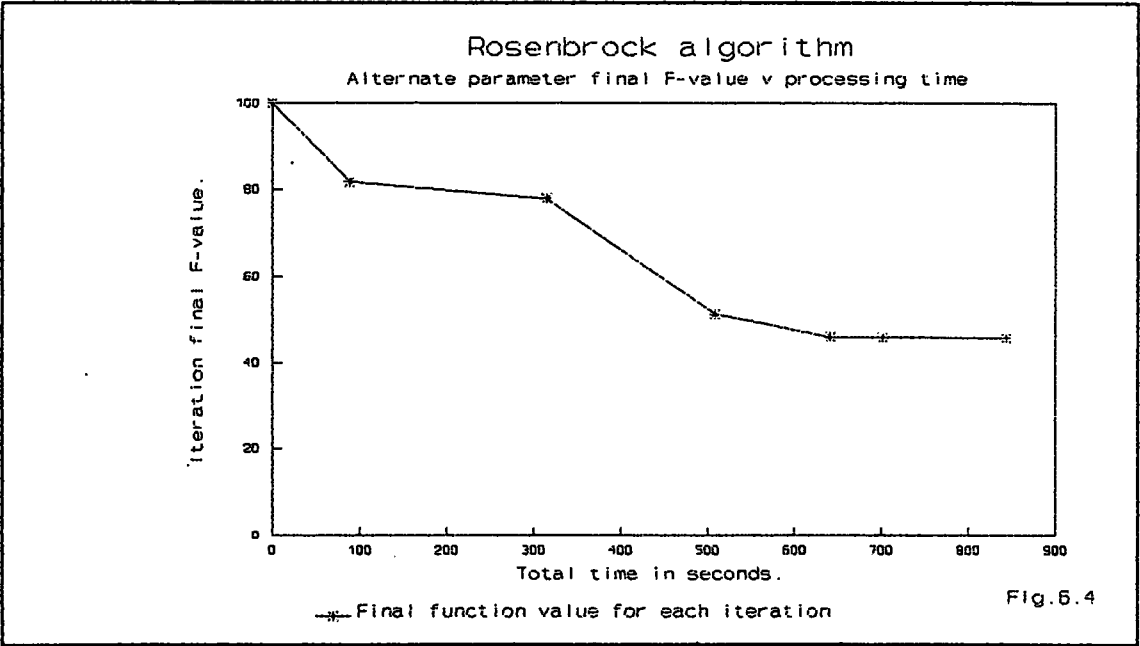
was changed by the author so as to examine each step on each axis in turn and then proceed to the next parameter axis. Tests were made using the same input data, rainfall and Penman open water evaporation index taken from Ray Catchment observations, and parameter starting point for each algorithm, and it was found that apparently the alternate parameter search was significantly less efficient at reducing the F-value than the alternate axes search.



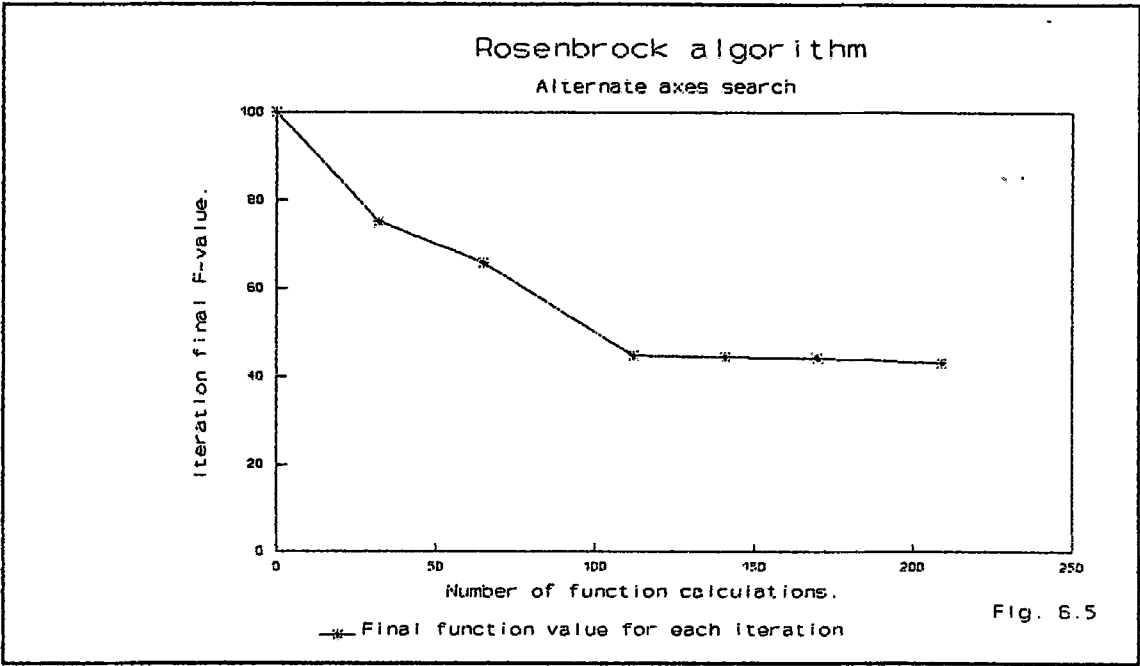
The relative position of the two methods is shown in Figure 6.2 where the alternate parameter search has a slightly higher F-value per model run at the end of the first iteration than that of the alternate axes search. The alternate axes search is more efficient than the other for iterations two and three, and has virtually found the optimal point by the fourth iteration. Both algorithms complete the optimization by the sixth iteration. The result is empirical since it was only tested on the data from one catchment.



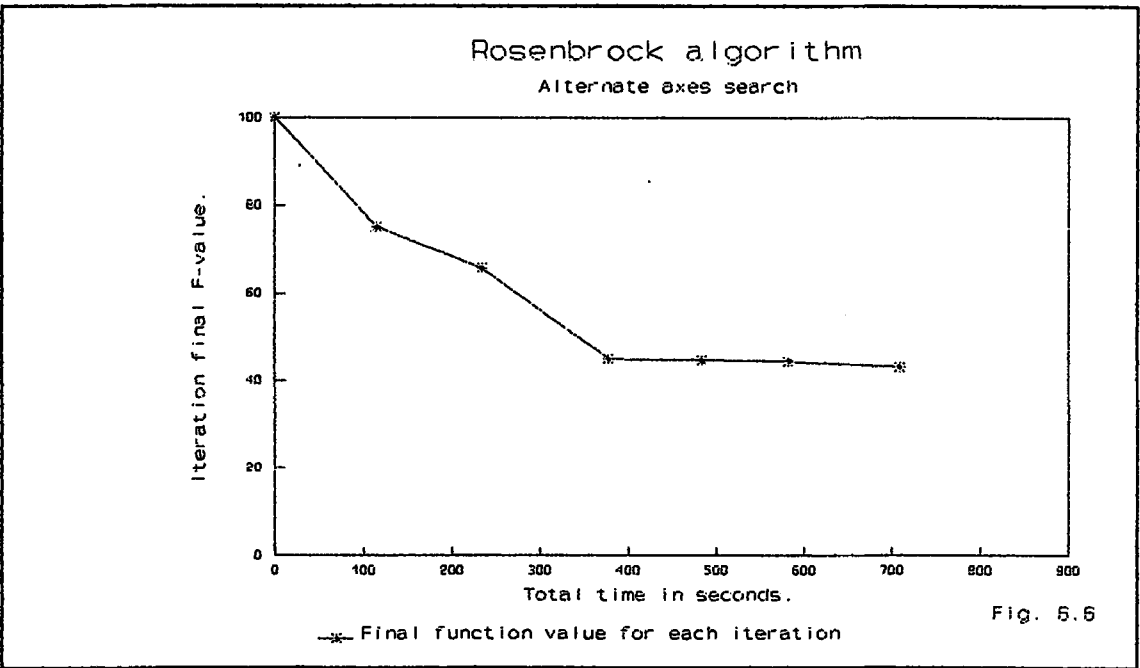
The plots of the final F-value for each iteration against the number of function calculations and total Central Processor



Unit (CPU) time used by the computer for the alternate parameter search are shown in Figures 6.3 and 6.4.



while the alternate axes search are shown in Figures 6.5 and 6.6.

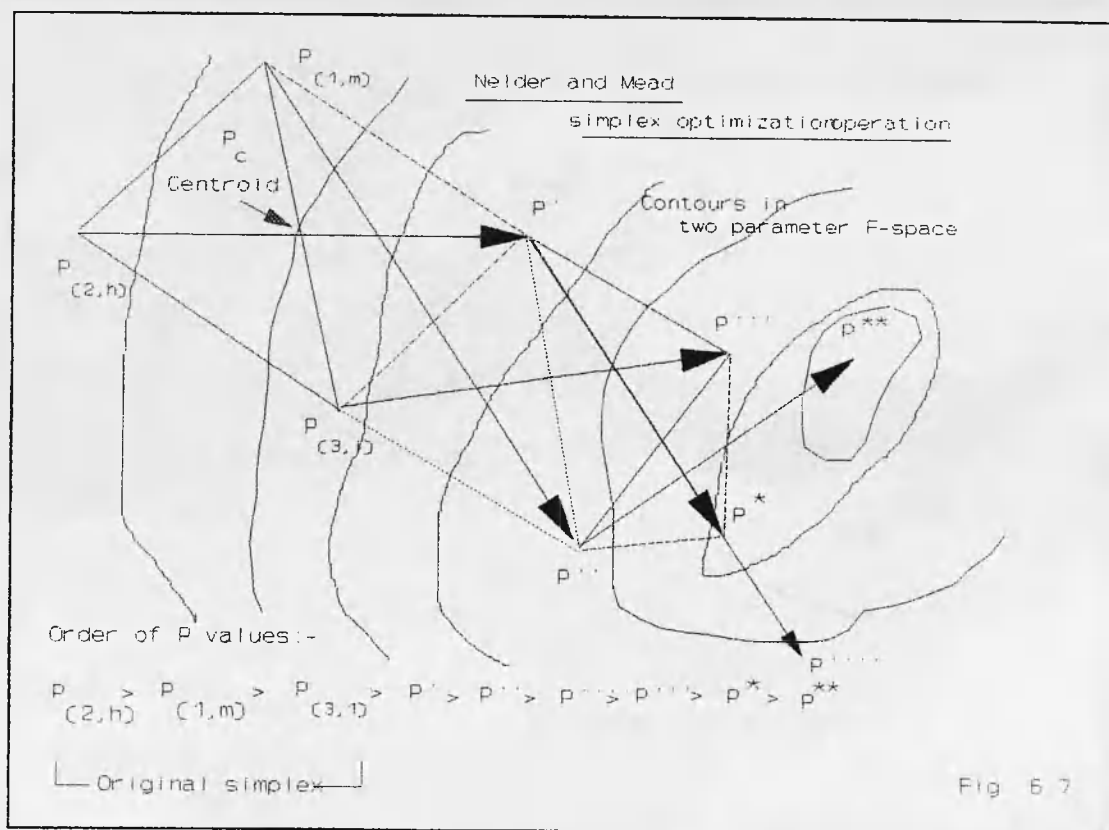


The similarity of the pattern between the number of function calculations and computing times lead to the conclusion that only one should be used in tests and that the former would be independent of the type of computer.

6.6 The Nelder and Mead Simplex algorithm.

This method is again one in which derivatives are not considered in the search of F-space. The technique was first published by Nelder and Mead (1965) but was not examined in the Ibbett thesis (1970). Pickup (1977) did test it and concluded that it was more efficient than the Rosenbrock algorithm and therefore recommended its use. The version used for this research is one which the author developed from that published by Clarke (1973). This published version unfortunately had a number of coding errors and the algorithm was inconsistent at certain points.

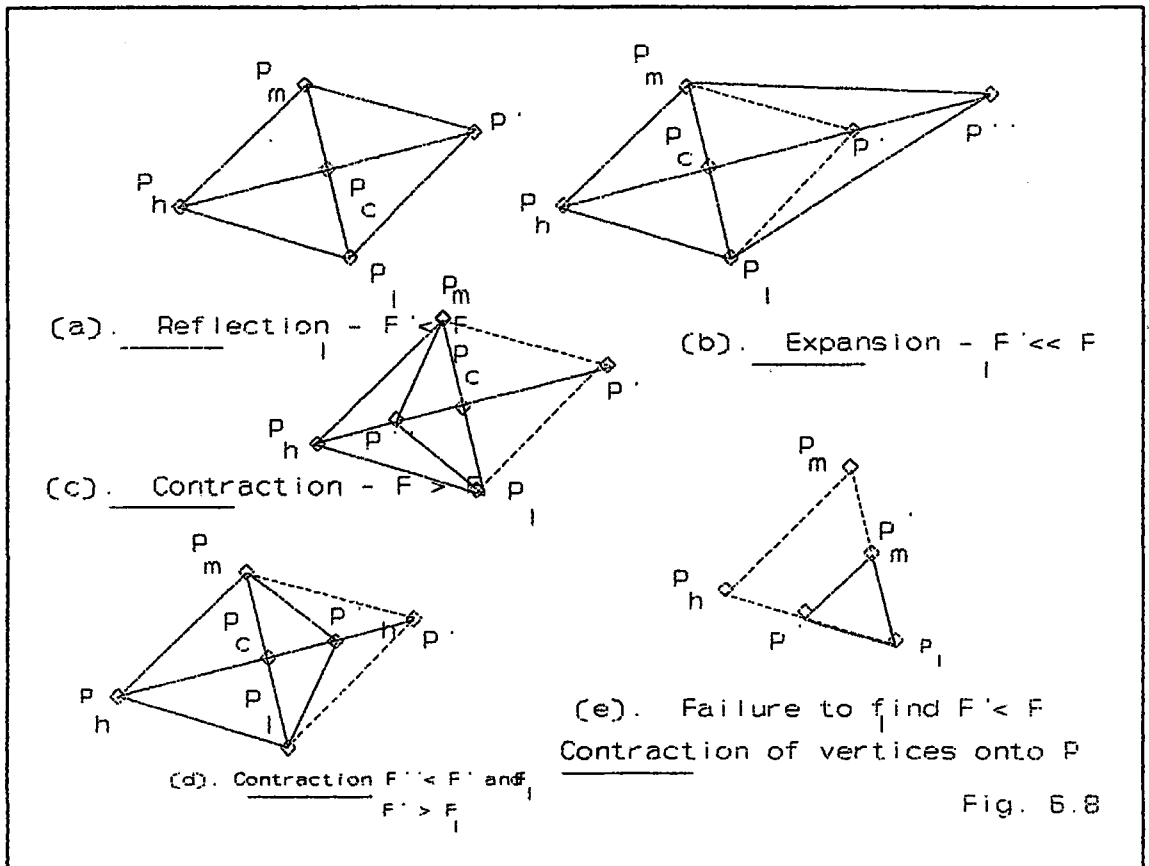
The Simplex algorithm searches for the optimum by comparing the points of a simplex which has $n+1$ vertices in the n -dimensional F-space spanned by the parameters. For example, a simplex in two dimensional space is a triangle and in three dimensions a tetrahedron. The space is searched by comparing the values of the objective function at the vertices, and replacement of the highest point in the simplex by another point. The general process is shown in Figure 6.7 for a two dimensional region with the original simplex defined by $P_{(1,m)}$, $P_{(2,h)}$ and $P_{(3,l)}$. Here the first integer denotes the order in which the parameter are 'stepped' to form a simplex: vertex 1 is the initial point while vertices 2 and 3 are the vertices obtained by stepping each of the two parameters away from initial value. The letters show the relative values of the objective function: $F_h > F_m > F_l$ at each of the vertices. The strategy adopted by the author after empirical tests was to allow the step to be either positive or negative alternately so as not to give a positive bias to the initial generation of the simplex. This assumes that the initial values of the



parameters are the best estimates available. The simplex is usually 'regular' as the 'step' size is fixed but Parkinson and Hutchinson (1972) showed that this is not a necessary condition for the algorithm to work.

The search is begun by the 'reflection' of the highest vertex, $P_{(2,h)}$, about the centre of the other two vertices to give the new point P' as shown by Figure 6.7. The value of the objective function at the lowest vertex, F_l , is then compared with the new value, F' , at the point P' . In the example given in Figure 6.7 $F(P')$ is lower than $F_{(3,l)}$. The value of F' value must satisfy one of the following three inequalities:-

$$\begin{aligned}
 (1). \quad & F_l \leq F' \leq F_h \\
 (2). \quad & F' < F_l \\
 (3). \quad & F' > F_h
 \end{aligned}
 \tag{6.7}$$



In the first case, (1), the highest vertex of the original simplex, $P_{(2,h)}$, is replaced by the new point, P' , and the process is then repeated using the simplex on $P_{(1,m)}$, $P_{(3,l)}$ and P' . This is shown in Figure 6.8(a).

For inequality, (2), the reflection is regarded as highly successful and the search vector is extended along the same line expanding the simplex by scalar multiplication with a coefficient, γ (>1), Figure 6.8(b). The reflection and expansion of the search vector is described by the vector relationships:-

$$\text{Reflection} \quad P' = (1+\alpha) \cdot P_c - \alpha \cdot P_h$$

$$\text{Expansion} \quad P'' = (1-\gamma) \cdot P_c + \gamma \cdot P'$$

where P'' is the vector extended from the previous reflected vector, P' , and P_c is the vector from the highest vertex to the

centroid of the remaining two vertices. The reflection coefficient is α , and the expansion coefficient is β .

The objective function value, F'' , at the new point is now calculated. If F'' satisfies $F'' < F_l$ then this means that the expansion of the simplex was successful, the old highest vertex is replaced in the set of points by the new point as a vertex and the process is repeated using the new redefined simplex. If $F'' > F'$ then P_h is replaced by P' and the search process continues with the reflection of the new simplex for the next iteration.

When the reflection of the highest point onto P' has $F' > F_h$ then it would appear that the search vector is moving away from a minimum, having passed over it. The simplex is therefore contracted by moving the point P' towards the assumed minimum L' so that the vector is equal to P_c reduced by a factor β (< 1). The vector relationship is:-

$$\text{Contraction } (F' > F_h) \quad P'' = (1-\beta) \cdot P_c + \beta \cdot P_h$$

$$\text{Contraction } (F_l < F' < F_h) \quad P'' = (1-\beta) \cdot P_c + \beta \cdot P'$$

These are shown in Figs. 6.8(c) and 6.8(d).

If $F'' < F'$ and F_l , and $F' > F_l$, then it is replaced by the new vertex at P'' . In the event of a failed contraction when a lower value of F has not been found then all the other vertices are contracted towards the lowest vertex and the search is restarted with the smaller simplex: the vertices are replaced by half the sum of each vertex and the lowest. This contraction of the simplex is shown in Figure 6.8(e). In the event of no further progress being made because of the contraction then a similar process expands the simplex by a factor of four to test a wider field.

The reflection coefficient α was introduced by Nelder and Mead but since they and subsequent workers found the best value of this to be unity it is not normally discussed. To obtain a true reflection the curvature of the topography of F-space would have to be accurately known and of a regular form if α is to be a constant. It is therefore much simpler to take α as 1 which one would expect from the normal plane geometry involved in a 'reflection'. Typical values for the other coefficients found by Nelder and Mead are $\beta = 0.25$, and $\gamma = 2.0$. Parkinson and Hutchinson (1972) obtained values of $\beta = 0.25$, and $\gamma = 2.5$, but were doubtful of their general applicability. General tests by the author showed that, in practice, the Nelder and Mead values were more efficient. The reflections of the sequence of different simplex leads to a 'zig-zag' approach to a minimum with the direction of the search vector approaching what can be a local optimum or a localized 'flat' condition of the vertices.

The problem remains of a suitable test for convergence of the algorithm, and the one currently in general use tests whether the simplex is lying 'flat' on the n-dimensional surface. After each iteration of the algorithm a test is made of the standard deviation, σ , of the $n+1$ objective function values, from each vertex:

$$\sigma = \sqrt{\sum_{i=1}^{i=n+1} \frac{(F_i - F_{\text{mean}})^2}{n}} \quad (6.8)$$

The search is terminated when σ falls below a preselected value. Bunday (1974) follows the scheme for this algorithm with a slight difference: he discards the new point found

because the value of F is too high instead of using the added information in contracting to form a new simplex. The test for convergence can be unsatisfactory since it is possible (though not common) to obtain this 'flat' condition without the simplex having contracted on to an optimum. Because of the small numbers involved in computing the standard deviation, double precision has to be used.

6.7 The Levenberg and Marquardt algorithm.

This is a method which requires the first derivatives of the objective function F in it's least squares form and is a modification of the Gauss-Newton algorithm suggested by Aaby and Dempster (1974). We compute

$$F = \sum_{i=1}^{i=m} (Q_o - Q_s)_i^2 \quad (6.9)$$

where m is the number of ordinates, F is the objective function value, Q_o is the observed stream flow and Q_s is the simulated flow.

Let $f_i = (Q_o - Q_s)_i$ then

$$F = \sum_{i=1}^{i=m} (f_i)^2 = \mathbf{f}^T \cdot \mathbf{f}$$

where \mathbf{f} is the vector of all the simulation residuals, f_i .

Differentiation of equation 6.9 with respect to the parameter p_j ($1 \leq j \leq n$) gives

$$\frac{\partial F}{\partial p_j} = 2 \cdot \sum_{i=1}^{i=m} f_i \frac{\partial f_i}{\partial p_j}.$$

This can be written in vector form as

$$\nabla F(\mathbf{p}) = 2 \cdot \mathbf{J}(\mathbf{p}) \cdot \mathbf{f}(\mathbf{p}) \quad (6.10)$$

The f_i are not linear functions of the parameters so then the Jacobian \mathbf{J} does not have a constant value and will depend on p_1, \dots, p_n . It has to be assumed that \mathbf{J} is quasi-constant as it varies slowly.

Therefore perturbing equation 6.10 to include the vector $\delta \mathbf{p}$

$$\nabla F(\mathbf{p} + \delta \mathbf{p}) = 2 \cdot \mathbf{J}(\mathbf{p}) \cdot \mathbf{f}(\mathbf{p} + \delta \mathbf{p}) \quad (6.11)$$

the vectors ∇F and \mathbf{f} are now functionally dependent upon the parameter vector, \mathbf{p} , and the vector $\delta \mathbf{p}$ represents the change in the vector \mathbf{p} to reach an optimum.

Expanding $\mathbf{f}_{(\mathbf{p} + \delta \mathbf{p})}$ to the first order term of the Taylor expansion and considering the second order and higher terms to be negligible

$$\mathbf{f}_{(\mathbf{p} + \delta \mathbf{p})} \approx (\mathbf{f} + \mathbf{J}^T \cdot \delta \mathbf{p})$$

$$\text{At an optimum} \quad \nabla F_{(\mathbf{p} + \delta \mathbf{p})} = 0 \approx 2 \cdot \mathbf{J} \cdot (\mathbf{f} + \mathbf{J}^T \cdot \delta \mathbf{p})$$

$$\text{Hence} \quad \mathbf{J} \mathbf{f} + \mathbf{J} \mathbf{J}^T \cdot \delta \mathbf{p} \approx 0$$

$$\text{so that} \quad \delta \mathbf{p} = -(\mathbf{J} \mathbf{J}^T)^{-1} \cdot \mathbf{J} \mathbf{f}$$

This analysis enables the optimum of a quadratic surface to be found. For non-quadratic surfaces the optimum correction

vector, δp , is used to define the vector search direction along which the line optimum is found, and this point is then used as the starting point of the next iteration. First derivatives are needed to set up the Jacobian J .

If $p + \delta p$ is an estimate for the optimal values of the parameters. The method can behave erratically if δp is not sufficiently small or the region of F -space is not sufficiently smooth, and random search directions may be generated. The optimum vector, δp , has to be bounded so that the modulus of $(p + \delta p)$ lies close to the starting point defined by p .

Levenberg (1944) first applied this technique for non-quadratic surfaces by applying a line search in the direction of δp . However, this often failed to produce a significant reduction in F , particularly in the initial stages. Marquardt (1965), proposed a method which modified Levenberg's search direction towards the direction of steepest descent and this proved to be very efficient. Fletcher (1980) examines the algorithm and discusses the computational difficulties.

6.8 NLFIT Package

This is an interesting suite of programs designed to determine and make inferences from the random errors associated with non-linear hydrologic models. The suite of programs was developed by Kuczera (1987); Version one whilst he was at Monash University and Version two at the University of Newcastle, N.S.W., Australia. NLFIT is the principle program which provides a 'shell' for the User's own model and uses the Marquardt algorithm to fit the model to observed input data. The program is specifically designed for non-linear models and

can handle errors with non-stationary variance and autocorrelation.

Two types of information can be used in fitting: several time series records of observed system responses, and prior information on the model parameters.

The principle output is a multivariate normal probability distribution for the model parameters which summarizes what is known about the parameters given the data used for fitting. The mean of this distribution gives optimized parameters and the covariance matrix provides measures of parameter sensitivity and interdependence. Further information is provided on model performance and error characteristics.

The package stands or falls on the efficiency of the Marquardt algorithm in fitting the operator's model parameters to the input data. This is in a similar fashion to the commercial HYRRROM package which is dependent on the Rosenbrock algorithm for calculating the optimal values of the parameters. However, the model cannot be varied. The models described in this thesis operate in a shell which allows both the model and the optimization algorithm to be varied by the operator. This opens up the possibility of an Artificial Intelligence package for hydrologic modelling provided that the ground rules for the use of such a system can be established.

The use of optimization algorithms is not always supported by experts in the field of hydrology. Early criticism of their use to find optimal values of model parameters in general was based on the loss of generality of a model by the modification of physically based equations when fitted to observed data by Dunin of Pilgrim and Dunin (1970). The errors in a model simulation were considered as important feedback to

enable the modification and improvement of the model concepts and the reduction of these errors by optimization therefore affected the validation of the model. Strangely enough the removal of errors and bias in observed data has also been objected to by some statisticians as a reduction in the 'random' error term in statistical analyses!

Chapter 7.

Algorithms new to the field of Hydrology.

7.1 The Numerical Libraries.

At the beginning of this research it was thought to be relevant to the work to use the compiled optimization algorithms available in software libraries. The most important software library used by Universities and large research organisations is the Numerical Algorithms Group (NAG) Library. Version 11 of this software was then available at the Institute of Hydrology mounted on their old IBM 4381 Main Frame running under the VMS operating system in 1984/85. The algorithms required a substantial amount of user effort to implement them: a selection of the algorithm had to be made, the appropriate Job Control Language mastered, and the values of parameters to fit the algorithm to the model and input data had to be chosen. An officer attached to the NERC Computing Services was there to be consulted at that time and to funnel queries and problems through to already successful users or as a last resort to the NAG organisation. From experience it would appear on average that a fortnight was needed to get an algorithm up and running from square one, and then further work was needed to ensure that the package parameters had been correctly identified and that sufficient workspace was available to contain the data generated. On the whole the most helpful of the options available at the time was the 'successful user' who has already struggled through the process. Unfortunately nobody at the Institute of Hydrology had tried to use the optimization section of the Library and there was therefore no previous experience available.

It would probably have been best to set up a simple 'dummy' model to test that the inputs to the NAG algorithm were understood correctly by the user and doing what they were said to do in the manual; at least one algorithm parameter did not appear to be working. This was discovered by the author during tests on the algorithm parameters using the nine parameter model.

Two algorithms were chosen from Chapter E04 - 'Minimizing or maximizing a function'. These were the 'easy-to-use' version of the Quasi-Newton algorithm, E04JAF, and the Conjugate direction algorithm, E04DBA/F. An outline of the theory for the latter is given in the next section, 7.2. The Quasi-Newton method approximates the Hessian matrix used in the Newton method, by a matrix which is modified at each iteration to include information about the curvature of the function at each point. This can be more efficient than Newton's method because the Hessian matrix need not be input or approximated by finite differences.

Daily data for 28 months were mounted as a modelling file from 7.64 to 12.66 for the Ray Catchment and the new nine parameter version of the model set up as FUNCT1.

The quasi-Newton algorithm took ninety seven calls to FUNCT1 (the model subroutine) for the four most 'sensitive' parameters, and completed the optimization by giving an under prediction from the model simulation on the total flow volume of 55%. Problems with the output from the model simulation statistics sub-routine were experienced which suggested program array overwriting and possible compilation mapping conflicts. There were no error diagnostics from these problems. The conjugate direction algorithm performed satisfactorily for the

same four parameters. It took 471 calls to FUNCT1, with a final error in simulated total discharge of -0.059%, and a correlation of 67%. A satisfactory performance at that time apart from the large number of model runs.

However, an attempt to repeat this success with the conjugate directions algorithm starting with the new optimal set of model parameters held steady, and another set of four parameters allowed to 'float' in order to improve the correlation, failed completely. Attempts to evaluate the diagnostic messages produced concluded that something had gone wrong within the NAG algorithm, due to inputs from the model. Attempts to diagnose these problems failed, and so the attempt to use the NAG Library was abandoned.

Only the experience of generating and compiling particular model codes for given problems will help with understanding diagnostic messages produced by the compiler. These usually have several depths of interpretation to arrive at the correct diagnosis. The use of 'Print' statements at various points in a program is the best method of solving this type of problem, or the running of a program under the control of a modern debugger such as the Microsoft Code View. However, these methods require access to all of the program code which was not made available to the Institute with the NAG Library due to the restrictions of the operating license purchased by the Institute. The program code can be obtained under license from NAG. For that matter an interactive model package such as HYRRROM does not have the program code available to the user. The latter simply rejects, with an appropriate (?) message, input data which is not suitable for processing by the package.

There are 'compile and run' libraries such as the Wiley Science Sub-routine Library and the Harwell Numerical Sub-routine Library in which the algorithm code is made available, and the whole program is compiled by the user. This enables the use of 'Print' statements or a debugger in the event of diagnostic problems. The Harwell Library has a large number of optimization algorithms, Section V, and two of these from Release 10, the Powell conjugate directions and Marquardt algorithms, have been used in this research.

7.2 The conjugate direction algorithm.

The method, published by Powell (1964), was one of the algorithms tested without success by Ibbitt (1970). The tests carried out by Ibbitt showed that the method was promising but had problems to which an immediate solution could not be found. Later work demonstrated that the step length used could take the search vector away from the local optimum towards another, and that the maximum number of iterations should be between 10 and 20, Fletcher (1972); this limit was later confirmed by tests made by Schwefel (1977).

The algorithm uses a variation of the alternating variables method such that when applied to a function of quadratic form it causes conjugate directions of search to be chosen, as defined by equation 7.1 below, and if applied to a general function the ultimate rate of convergence should be quadratic.

The directions p_i and p_j are said to be conjugate with respect to the positive definite matrix G if

$$(p_i)^T \cdot G \cdot p_j = 0 \quad (7.1)$$

It can be shown from this definition that if $p_0, p_1, p_2, \dots, p_{n-1}$ are n mutually conjugate directions in n -dimensional space then they are linearly independent, i.e.

$$\sum_{i=0}^{n-1} \alpha_i \cdot p_i = 0 \quad (7.2)$$

holds only if all the α_i are zero; if the matrix G is replaced by the unit matrix I then the p_i are mutually orthogonal.

The strategy of conjugate directions is based on the concept that a line through the minimum of a quadratic objective function cuts all contours at the same angle. Powell (1964) uses this idea to construct such directions for a sequence of linear searches. For an objective function, F , of n variables the unit vectors are taken as initial directions for the first n line searches, and then a line minimization is carried out using Powell's quadratic interpolation in the resultant search direction. One of the old direction vectors is then replaced by this resultant vector. The discarded search directions p_m , corresponds to the value of Δ such that

$$\Delta = \max_{1 \leq m \leq n} (F_{p_{m-1}} - F_{p_m})$$

where L_{m-1} and L_m are successive points.

The old directions are retained if the function for the i 'th iteration $F_{i+2} \geq F_i$ and/or from Powell's interpolation

$$(F_i - 2F_{i+1} + F_{i+2}) \cdot (F_i - F_{i+1} - \Delta)^2 \geq \frac{1}{2} \cdot \Delta \cdot (F_i - F_{i+2})^2$$

where F_i, F_{i+1} and F_{i+2} are the last three evaluations of F .

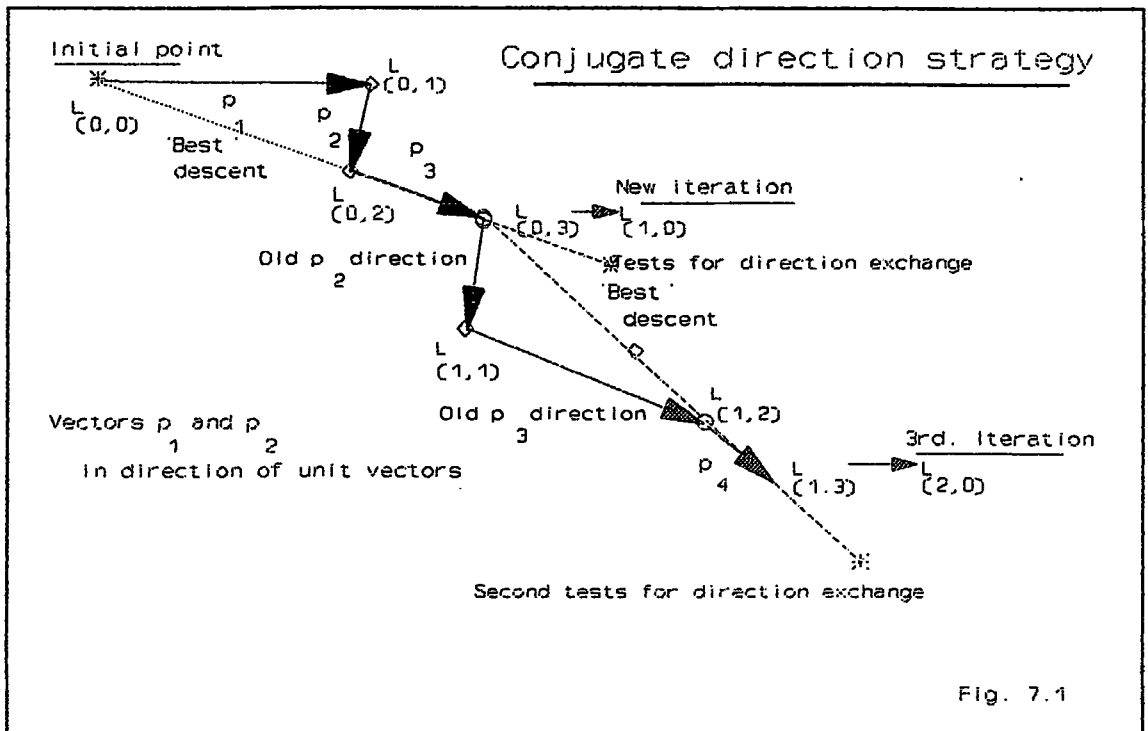
The final point of that iteration is then used as the initial point for the next iteration. Powell demonstrates that, provided the objective function is quadratic and the line searches are exact, after n cycles consisting of $n+1$ line searches, a set of conjugate directions is obtained.

In this special case

$$G = \nabla^2 F(p) \quad (7.3)$$

the Hessian matrix, and the minimum of a quadratic function is obtained exactly in n line searches in the directions p_j . For general non-linear problems the convergence rate cannot be specified. It has to be assumed that the problem's objective function behaves approximately as a quadratic around a local optimum if the conjugate directions algorithm is to be employed. Powell's criterion for convergence if a point is reached in an iteration of the algorithm where each of the variables has changed by less than $0.1 e_k$, where e_k is the required accuracy for the k -th element of the solution. These tolerances are chosen so that they are roughly proportional to the expected values of the final p_k .

The process is illustrated for a two parameter case in Figure 7.1 - the initial point $L_{(0,0)}$ is chosen together with the required accuracy ϵ for the parameters. The unit vectors are found and used as initial directions for the first two line searches, p_1 and p_2 , giving the points $L_{(0,1)}$ and $L_{(0,2)}$. A line search is then carried out in the resultant direction of 'best' descent, p_3 , which produces point $L_{(0,3)}$. Tests for direction exchange are made, and suppose that this leads to the retention of the p_2 and p_3 search direction vectors. This concludes the



line searches for the first iteration case, $n = 2$ (two parameters), and $n+1 = 3$ line searches; the first of the old direction vectors is eliminated, the subscript of the remaining one is reduced by 1 and the new vector takes the place vacated. For the starting point of the new iteration $L_{(0,3)}$ becomes $L_{(1,0)}$. Then line searches along the old directions p_2 gives point $L_{(1,1)}$, and p_3 the point $L_{(1,2)}$ take place defining a resultant direction from $L_{(1,0)}$ to $L_{(1,2)}$ along which the search for the minimum produces the vector p_4 at the point $L_{(1,3)}$. A second set of direction exchange tests based on the value of F along the resultant vector is made, and the second iteration comes to an end with the point $L_{(1,3)}$ becoming the point of origin of the third iteration $L_{(2,0)}$. The process continues with either the old directions or a new one added in place of an old direction. This continues until the final point is found which satisfies the criteria for convergence: the change in each parameter is less than $0.1\epsilon_k$.

Powell uses quadratic interpolation for the line searches with the above convergence criterion to terminate the line search. This is also used as the minimization function in Chapter 6, Section 6.4, on the Rosenbrock algorithm when the section of the axis identified by the line search as containing the minimum has been found. The values of the three points taken to estimate the new minimum point are L_0 , $(L_0 + s_1 p_1)$, and either $(L_0 + 2s_1 p_1)$ or $(L_0 - s_1 p_1)$ depending on whether $F(L_0 + s_1 p_1)$ is less than $F(L_0)$ or not. The initial value of the step length s_1 is given by $e_1 \cdot s_{\max}$ as a proportion of the maximum step length and this changes during the operation of the algorithm. For the subsequent changes Powell uses the empirical recursion formula

$$(s_1)_k = 0.4 \cdot \sqrt{F(P_1)_k - F(P_1)_{k-1}} \quad (7.4)$$

With three points a , b , and c , where $a < b < c$, and the objective function values F_a , F_b and F_c the minimum can be estimated together with the second derivative of the quadratic function $F(L)$. The equation for the curvature ρ_1 (assumed positive) in the direction of p_1 is:-

$$\rho_1 = \frac{\partial^2}{\partial s^2} F(P_0 + s_1 \cdot p_1) \quad (7.5)$$

$$= \frac{(b-c)F_a + (c-a)F_b + (a-b)F_c}{(b-c) \cdot (c-a) \cdot (a-b)}$$

Powell re-uses p_1 for all the following interpolations in the direction p_1 , which is a gain in computing time as each direction is used several times. The predicted minimum point, L^* , is then:-

$$L^* = L_0 + \left[\frac{1}{2} s_1 - \frac{1}{p_1 s_1} (F(L_0 + s_1 \cdot p_1) - F(L_0)) \right] \cdot p_1 \quad (7.6)$$

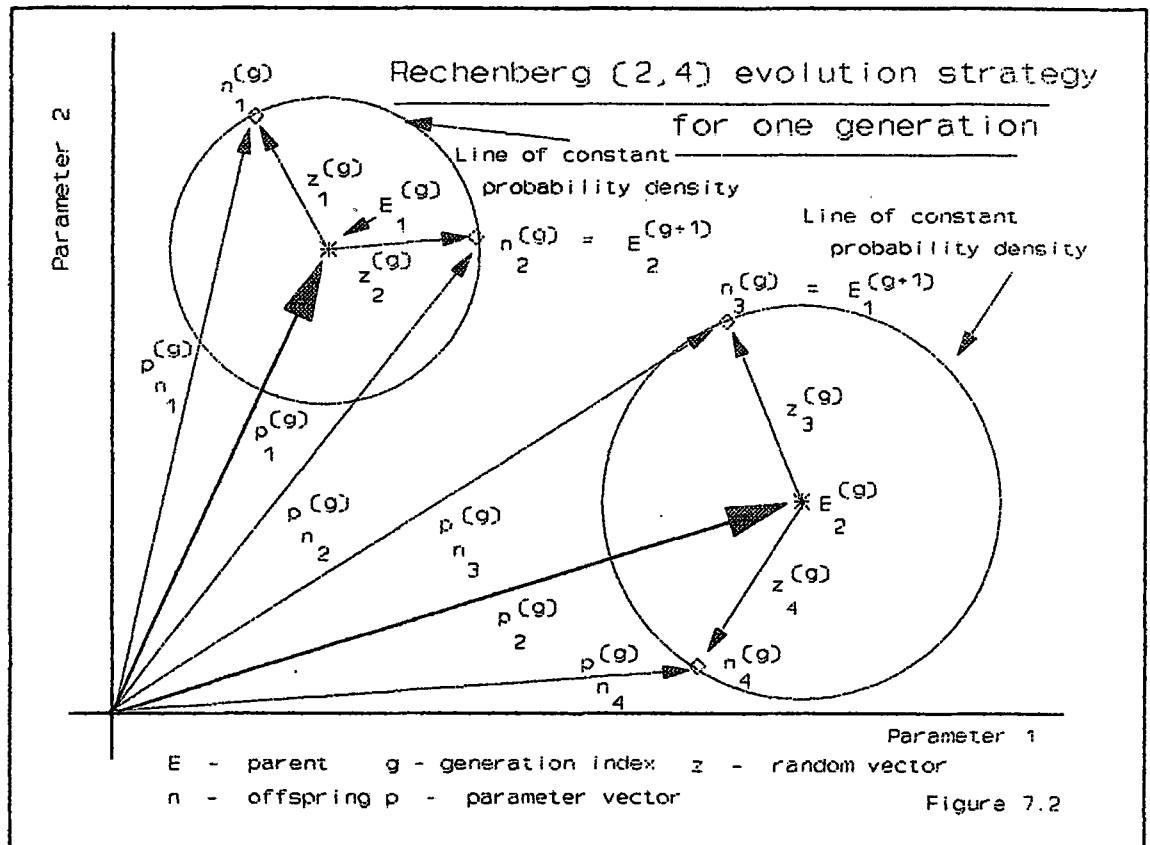
The Fortran program code by Powell is given in the Harwell Subroutine library as subroutine VA04A. In the specification to Harwell subroutine VA04A he mentions the problem of the step length and the necessity of limiting it to prevent different local optima being found. In Powell's 1964 paper two problems are discussed: new directions are chosen less often as the number of parameters increases; as the step lengths tend to zero the interpolation formulas become unstable.

7.3 The Evolution Strategy algorithm of Rechenberg.

This algorithm is a very unusual one in that it employs a random search technique structured according to the ideas of the evolutionary theory of Darwin and Wallace using natural selection. Whatever the problems associated with the theory of evolution, the apparent use of a naturally occurring 'optimal' strategy in nature provides some further justification for the logic of the theory. However, Anti-Chaos Theory is now showing, or postulating, that the original theory is too complicated an explanation of the origin of species. The original evolution concepts were used in the method employed as a basis for the algorithm, and this was first given as a lecture at the annual conference of the WGLR at Berlin in September 1964. A rough draft of the proceedings was translated by Toms of the RAE, Farnborough, in August 1965, as Library Translation No. 1122.

The method was later published as a fully developed algorithm in the book by Rechenberg (1973), and the program

code for the algorithm was given by Schwefel (1981) together with an analysis of the method and comparison with other algorithms working on theoretical problems. The use of an algorithm based on biological theory and observations would seem to be particularly appropriate for use on practical problems concerned with the natural environment.



The optimal scheme is based on the simulation of biological evolution using the combined ideas of mutation producing a small difference in the 'offspring', $n^{(g)}$, from the 'parent', $E^{(g)}$, and the survival of the former depending on some criterion for that generation, (g) . The way in which this strategy is modified by Schwefel and put into operation in the multi-membered case (μ, λ) is shown in Figure 7.2; μ is the number of parents and λ is the number of their offspring which are 2 and 4 respectively for the diagram. The population rules

in biological terms for this modified form of the algorithm are as follows:-

1. The population consists of μ individuals each of which are characterised by their genotypes consisting of n genes, which completely determine their ability for survival.
2. Each individual parent produces λ/μ offspring so that a total of λ new individuals whose genotypes are slightly different are produced by random mutation.
3. Only the μ 'best' of the offspring survive to become parents of the next generation to maintain the population at the original size.

In mathematical terms this becomes:-

Let the generation index $g = 0$ at the start of the iterative scheme. By choosing values for the parameters we can define an initial set of parents $L_k^{(0)}$, $k = 1(1)\lambda$ such that each parent satisfies any necessary constraints.

A new generation is spawned:-

$$\begin{aligned}
 n^{(g)} &= p_k^{(g)} + z_j^{(g)} \\
 \text{where } j &= (k-1) \frac{\lambda}{\mu} + \varepsilon \\
 \text{for } \varepsilon &= 1(1) \frac{\lambda}{\mu} \\
 k &= 1(1) \mu
 \end{aligned}
 \tag{7.7}$$

Here $z_j^{(g)}$ is a normally distributed random vector with mean zero and standard deviation $\sigma^{(g)}$.

For each of these off-spring we reject those which do not fall within the feasible region. Of those which remain we define the parents for the $g+1$ th generation as

$$p_k^{(g+1)} = \begin{cases} n_j^{(g)} & \text{such that } F(n_j^{(g)}) < F(n_k^{(g)}) \\ & \text{for } j = (k-1) \frac{\lambda}{\mu} + e, \quad e = 1(1) \frac{\lambda}{\mu} \\ p_k^{(g)} & \text{if } F(p_k^{(g)}) < F(n_1^{(g)}) \end{cases} \quad (7.8)$$

The generation index is then increased from (g) to $(g+1)$ and the process is repeated.

The offspring in Figure 7.2, $n_2^{(g)}$ and $n_3^{(g)}$, are chosen as the parents for the next generation because the F-values for these two points are less than their respective parents' values. In the (μ, λ) strategy with large values of λ there is less chance of a few or none of the offspring being 'better' than the parents, but it is possible for all descendants to be 'lethal' mutations and therefore the population to be in danger of extinction. In order to continue the optimization process the parents have to be allowed to survive to produce a new generation where the generation index (g) changes to $(g+1)$ and then to $(g+2)$.

The number of descendants per parent is limited by computer resources but it is intuitively obvious that too large a number would be self-defeating in making too precise a search around the parent or over-lapping directions, and thus slowing the rate of convergence.

The step lengths and random vectors, \mathbf{z} , are defined by random mutation of the vector components derived from the normal distribution $(0, \sigma^{(g)})$. The optimal variances, for an n -parameter model,

$$\sigma \propto \frac{1}{\sqrt{n}}$$

for the parameters $(p_{e,i}; i = 1(1)n)$

There are a set of strategy parameters, $\sigma_{e,i}$, to each parent E that describe the variances of the parameter random changes. Each descendent, n_e , of the parent E should differ from it in both $p_{e,i}$ and $\sigma_{e,i}$. These changes in variances should also be random, small and with expectation value zero.

Whether a descendant becomes a parent depends on $F(p_{e,i})$; which values of the variables, or model parameters, it represents depends not only on the F -value of the parents, but also on the standard deviation $\sigma_{e,i}$ which affects the size of the changes in terms of the position vectors of descendant and parent:-

$$z_i^{(g)} = p_{e,i}^{(g)} - p_{z,i}^{(g)} \quad (7.10)$$

In this way the step length also plays an indirect role in the 'selection' process. The convergence of the evolutionary process is defined either absolutely or relatively by the closeness of each parent's F -value in the generation. From the population of μ parents $[Ek; k=1(1)\mu]$ let F^* be the best function value:

$$F^* = \min_k [F(p_k^{(g)}); k=1(1)\mu]$$

and F^+ the highest in the generation

$$F^+ = \max_k [F(p_k^{(g)}); k = 1(1)\mu]$$

then for convergence

$$F^* - F^{**} \leq \varepsilon_c$$

$$\text{or } (F^* - F^{**}) \frac{\mu}{\varepsilon_d} \leq \left| \sum_{k=1}^{\mu} F(p_k^{(g)}) \right| \quad (7.11)$$

where $\varepsilon_c > 0$ and ε_d depending whether this expression is affected by the size of the computer precision. It is possible that descendants from parents could be converging on different local optima in which case the local optimum accepted by the algorithm would depend on the lowest value of F in a generation.

The computer program code for the algorithm as given in Appendix 2.3 of Schwefel (1981) was rewritten in FORTRAN77 with additional modifications to the code made necessary by the Microsoft Compiler Version 5.1. It was first tested on the Rosenbrock Parabolic Valley with $\mu = 2$ parents and $\lambda = 6$ descendants and found the minimum at $x_1 = 1.0$, $x_2 = 1.0$ and $F(x_1, x_2) = 0.0$.

The schema for the coded algorithm is given in Figure 7.3 with the control program for the input of parameters and specification of input/output files called **EVOL**. This main program calls the subroutine **KORR** which controls the optimization process, and then calls the **MODEL** subroutine when the optimized parameters have been returned to it from **KORR** for validation and generation of the final simulated data. The optimization subroutine is the multi-membered (μ, λ) version of the Rechenberg (1,1) algorithm developed by Schwefel. The constraints vector is stored in **CONSTR** by **EVOL** and accessed by **KORR** as required; initially the input parameter vector elements are checked in **CONSTR** to see that they lie in feasible space.

The sub-routine **OUTERR** does the initial validation of the optimization algorithm parameters and either accepts or

modifies them, and returns a warning flag to KORR if there is a significant error. The details of this process are output to channel 11 previously defined as the error file in EVOL.

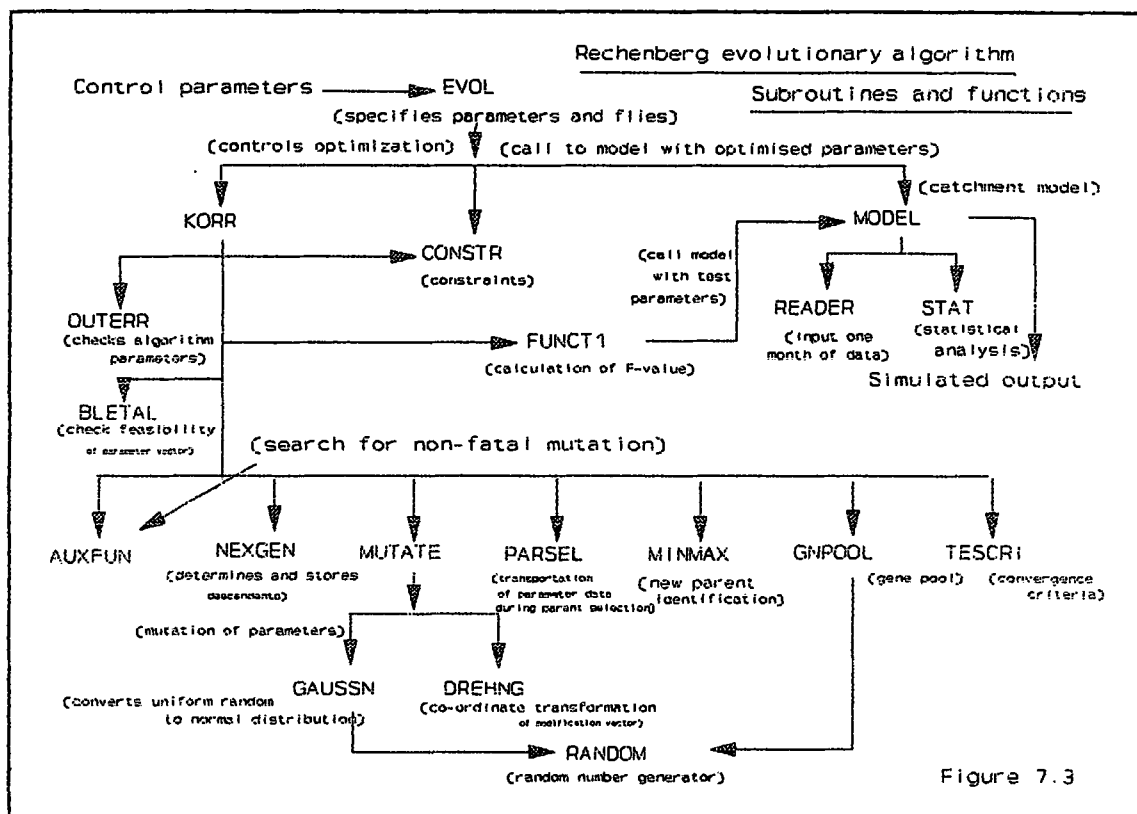


Figure 7.3

BLETAL is a logical function which checks the feasibility of the vector of parameters in relation to the constraints vector and flags the new parameter vector name to **KORR** as being 'lethal' if constraints are violated. **KORR** in turn calls **AUXFUN** which searches for a new vector lying within the feasible space using an auxiliary function to reduce the size of parameter 'steps' until this state is achieved.

FUNCT1 is called from **KORR** when a new parameter vector has been found. **FUNCT1** then calls **MODEL** to calculate the F-value and returns this value to **KORR** when obtained. **MODEL** calls **READER** which reads in and overlays the data arrays with the current month of rainfall and evaporation data for dynamic

processing month by month. When MODEL is producing the data for an optimal simulation called by EVOL it calls STAT each month in order to produce a statistical analysis for the complete run of simulation data.

NEXGEN identifies the best offspring, and transfers the data of a descendant representing a successful mutation to the data pool of parents for the next generation. The random alteration of the step sizes and parameters is handled by MUTATE by changing the DELTA and DELTAI parameters by multiplication by a random factor with a log-normal distribution. Mutate calls GAUSSN which in turn calls the new function TIMEDEL to produce a random seed based on one hundredths of a second from the computer clock, Appendix 3. This random seed is then applied to the Microsoft compiler functions SEED and RANDOM, and this latter sub-routine generates a uniform random number distribution in the range (0,1). The output from this is converted by GAUSSN to an element from a log-normal distribution; the function has been programmed to use the trapezium algorithm by Ahrens and Dieter (1972). MUTATE also calls DREHNG to perform the transformation of the co-ordinates of the modification vector for the model parameters.

PARSEL is called when the descendants are not an improvement in reduction of the F-value compared to the parents, and they have to be subject to selection again in order to produce offspring for the second generation. The sub-routine ensures the preservation and transportation of parent parameter data during selection for the next generation. MINMAX identifies the descendants with the minimum F-value or the worst F-value in order that TESCRI can apply the

convergence criterion. MINMAX also selects the order of descendants for survival. The sub-routine GNPOOL is called once for each of the parameters and their related step sizes. It supplies a set of parameters for a descendant by drawing on the pool of all the parents' genes in accordance with the type of recombination selected. It calls on the uniform random number generator RANDOM to generate the (0,1) changes in step sizes from the standard deviation of the changes already tested. When there is an intermediate recombination for the new positional vector angle the differences between the angles of the parent positional vectors are used to determine a suitable mean vector.

The convergence criteria are tested by TESTCRI after each mutation according to the preselected difference in F-values for the parameter vector, and the result transmitted to KORR to stop the search if convergence has been obtained. When this happens an optimal set of parameters is then passed to EVOL which calls for the simulated output data to be generated by MODEL.

Due to the problems previously demonstrated in Chapter 6 with computer timing in relative CPU and I/O times no sub-routine was used to account for this variable as previously used with this algorithm. Instead, the number of model iterations was used to limit the time in which the EVOL algorithm was running.

Chapter 8.

Data used for algorithm tests and comparisons.

8.1 Catchment data and preliminary analyses.

The importance of good unbiased data as an essential part of the modelling process has already been emphasised in the introduction. With good quality control over the data acquisition process, bias and substantial instrument error should not be present in the data set. However, in general good quality control of hydrological data is not consistently applied, and the overall run of data not often analyzed as fresh data is added to the set. It is therefore best before using a particular data set as a basis for simulation to see if it is homogenous and consistent without gross errors or bias. The tests developed by the author to establish whether these conditions are present or not are used in the later part of this chapter on the three sets of catchment data.

The data sets selected for the modelling process are rainfall, streamflow and the basis of the hydrological system 'loss' function, Penman open water evaporation. The three catchments from which these sets are taken as observed data have been chosen from the Kirkton partially forested catchment at Balquhiddy in Scotland, the headwaters of the Ray River in Buckinghamshire, and from the montane rain forest on the Lagan River near Lake Victoria in Kenya. These data sets are not specially selected for a high degree of correlation with output from the models, but as examples of different lengths of data from a wide altitude range, different rainfall and evaporation regimes, and different vegetation covers.

8.2 Data from the Kirkton catchment, Balquhiddar, Scotland.

The Kirkton catchment data is taken from the two experimental catchment areas which were set up to determine the differences in water use and sediment yield between the Kirkton with 44% under mature pine forest, and the Monachyle under a mixed heather, grass and bracken cover. This part of the experiment was due to run until early 1986, when staged clear felling would begin in the Kirkton, and drainage and initial planting in the Monachyle would be expected to change the initial results, and hopefully give data on land use changes which could then be extrapolated to the rest of upland Scotland.

The Kirkton data was chosen for a comparison set with the lowland English clay catchment and the montane rain forest catchment in Kenya. The comparison between the Kirkton and Monachyle catchments has already studied by Eeles and Blackie (1993). The Kirkton catchment has the change in landuse mentioned above to which the large landuse model can be applied for a comparison between the two models.

The Kirkton is 685 ha in area and has an altitude range of 250-850 m. Both catchments are steep sided glaciated valleys aligned approximately with their headwaters to the North and outfalls to the South. Their soils are peats, peaty gleys and upland brown earths. These overlay mica-schists and variable depths of glacial debris in the bottom of the valleys. The underlying geology is mainly Ben Lui Schists and there is an outcropping of Loch Tay Series of metamorphosed limestone in the Kirkton.

Rainfall on the Kirkton catchment is estimated using a network of eleven period gauges and a final total of three

automatic weather stations at different altitudes. During periods of snow cover the precipitation is estimated using snow gauges mounted at five accessible sites and the catchment mean estimated by regressions between the data from these sites and the complete network of rainfall gauges. The daily average rainfall is calculated from weighted domain areas for each gauge. Stream flow is measured by a Crump Weir for which the rating has been adjusted after intensive current metering.

Daily data from May 1984 to December 1988 has been used in the modelling work for rainfall and flow, with Penman potential open water data obtained from a weighted mean of the high and low weather stations in the Kirkton. The start of the calibration data in 1984 in May was chosen because this was the earliest period which avoided the complication of snow in the winter of 1983/84, and has a marked recession in the streamflow necessary for the calculation of the initial groundwater storage.

The monthly and annual totals of this data for the years 1984 to 1988 are given in Table 8.1 with Tables to 8.7 at the end of this Chapter. Mean and standard deviation figures are shown for the complete years. Rain days are the number of days on which rain fell with gauge readings from 'Trace' upwards. The difference in monthly rainfall and flow, $R-Q$, is shown as a crude guide to changes in soil moisture and groundwater storage and the actual evapotranspiration loss during the month. The negative figures are an indication of the drainage delay for a large storm or snowmelt event. Table 8.2 gives the frequency distributions of the three sets of data on a volume basis percent and time in days percent. The intervals and

their sizes are standardised for the three catchments for comparison.

8.3 Data from the Ray catchment, Grendon Underwood, Buckinghamshire.

This experimental catchment in Buckinghamshire was the first set up by the Institute of Hydrology to obtain data for the development of flood prediction models for lowland clay catchments.

The Ray catchment has an area of 1856 ha with an altitude range 50-187 m. The River Ray is a tributary of the River Thames, and the catchment is underlain by Jurassic Clays which are mainly Ampthill and Oxford Clays. The highest point of the catchment is Quainton Hill which forms a limestone outlier from the Chilterns.

Movement of water into the soil is through cracks or root paths in the clays. The presence of these clays giving an apparently watertight catchment was the original reason for the experiment being set up, but the hills that form the northern watershed have a clay capping with an underlying thin covering of glacial drift of unknown area which forms a small trapped aquifer. There is some further glacial head towards the centre of the catchment, and the southern boundary is free from these deposits. In dry summers the stream stops flowing, and this ephemeral nature of the flows make it quite difficult to model the low flows from the outfall of the catchment.

The land use of the catchment in 1981 was arable - 40%, grassland - 45%, and woodland - 15%. The data used runs from January 1964 to December 1981 which is the most viable and error free sub-set of the whole time series. Data are available from 1982 onwards but this has not been assessed for

reliability and bias. The starting month was again chosen for the uncomplicated recession of the streamflow.

The raingauge network consisted of twenty sites with standard meteorological gauges whose distribution depended on volunteer observers, and three equipped with Dines recording raingauges (later reduced to one). There were two further recording gauges at each of the meteorological sites at Grendon Underwood Prison and Quainton Hill. An automatic weather station was operated from 1976 together with the manual meteorological site at the Prison.

Daily rainfall for the catchment was obtained from this network by means of Thiessen polygon weighting for each gauge. The Penman Open Water index came from the meteorological sites, and the stream flows were estimated from a critical depth trapezoidal flume at the outfall. The monthly data and statistics for the years 1964 to 1980 are shown in Table 8.3 and daily frequency distributions in Table 8.4. Descriptions of the Tables and their statistics are the same as at the end of section 8.2.

8.4 Data from the Lagan River montane rain forest catchment, Kericho, Kenya.

This catchment is on the edge of the escarpment above Lake Victoria near the town of Kericho, and was to form the control catchment for the tea estate catchment excised from the montane rain forest some two kilometres to the east. Unfortunately, its proximity to Lake Victoria affected the rainfall regime and it could not be used as a control. The land use question, for which the two catchments were set up in 1957 to answer, was whether the indigenous forest can be replaced by an

economically valuable crop without damage to soil and water resources.

The catchment is situated in the south west Mau forest reserve in Western Kenya and lies within the drainage basin of Lake Victoria. The catchment area of 543.7 ha is within 0.5° South of the equator with a mean altitude of 2,200 m, and the slopes of the catchment rise to about 600 m above the outfall.

The area is on fissure free phonolite lavas which are remarkably uniform in composition, and have weathered at the surface into deep stonefree soils, heavily leached and uniform in physical structure to a depth greater than 6 m with a high infiltration capacity.

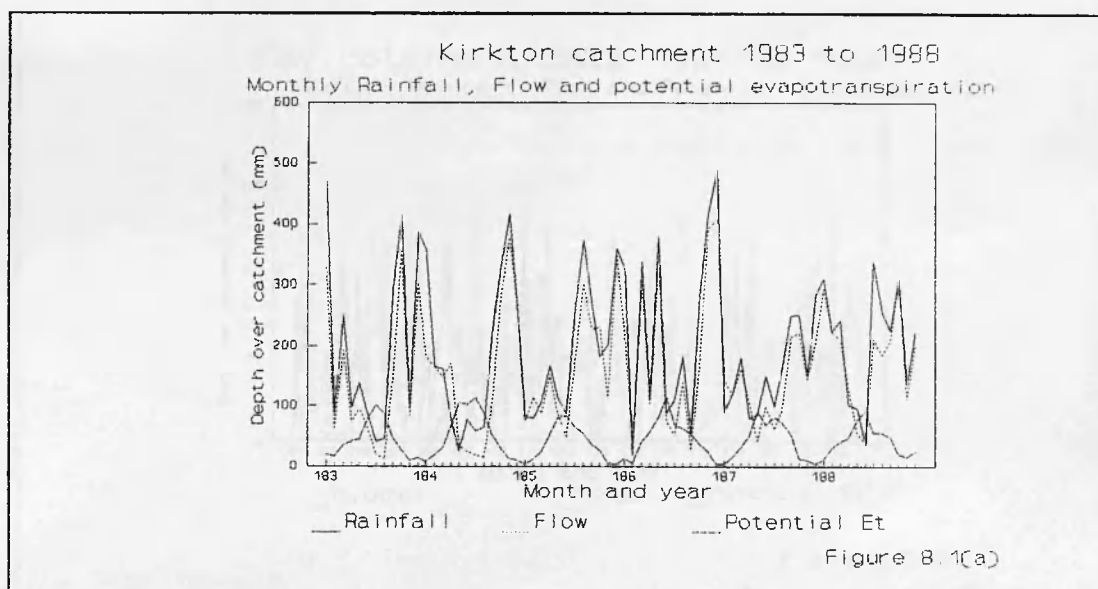
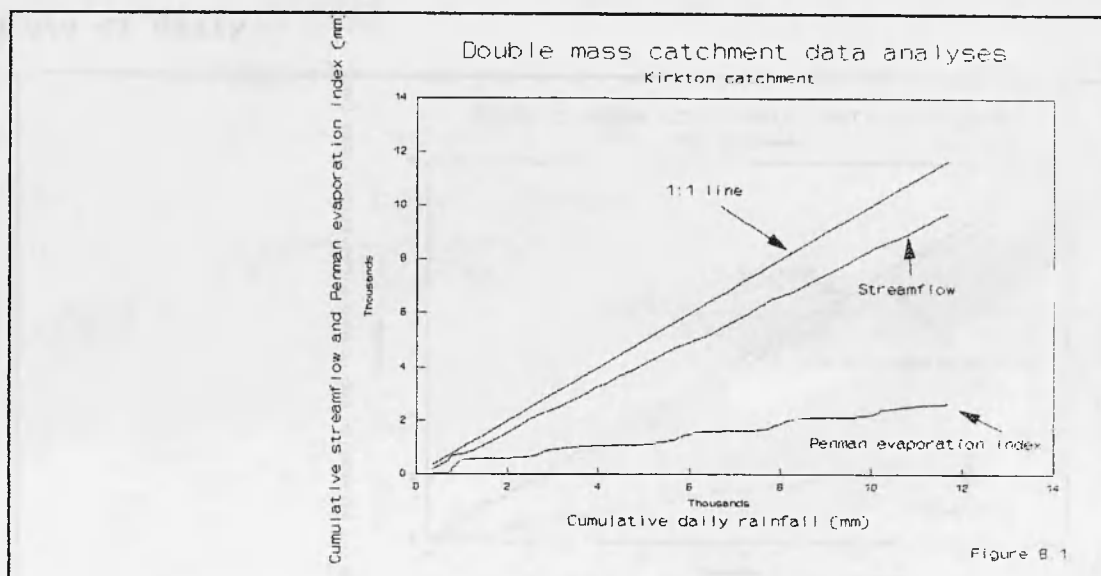
The rainfall was measured by two Dines recording gauges in large clearings on the south of the catchment, and three daily standard raingauges - two with the recording gauges and one mounted on a post in a clearing between the other two. Estimates of the catchment mean daily rainfall were made using Thiessen polygon weighting for each standard gauge. The meteorological data was obtained from a site at the Tea Research Institute about 2 Km to the North. Streamflow was gauged by a compound sharp and broad crested weir.

The monthly and annual data totals for 1958 to 1973 are given in Table 8.5 with the daily frequency distributions in Table 8.6.

8.5 Double Mass analyses of data.

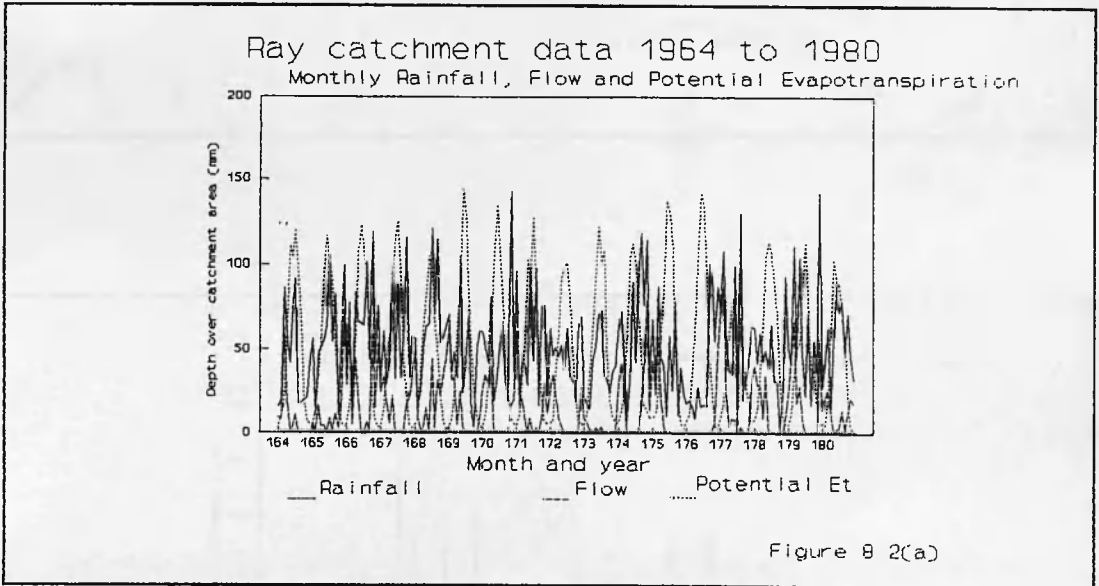
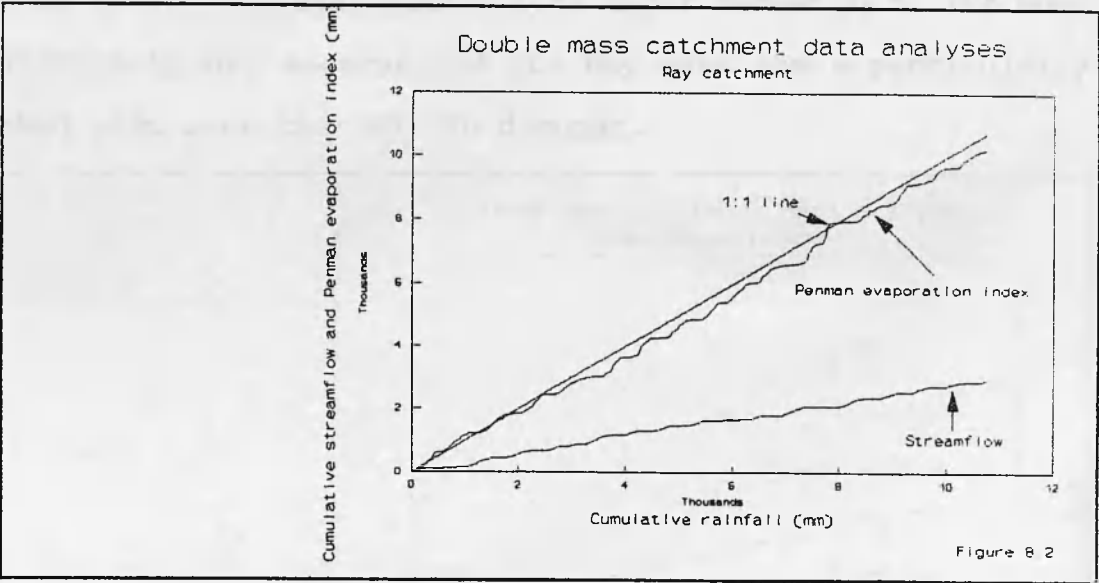
It is necessary to establish whether the data is homogenous over the time series, and has no problems due to changes of instruments, installation position or exposure. The best way of testing for these problems is to use double mass

curves: cumulative plots of the data sets against each other, or against time.



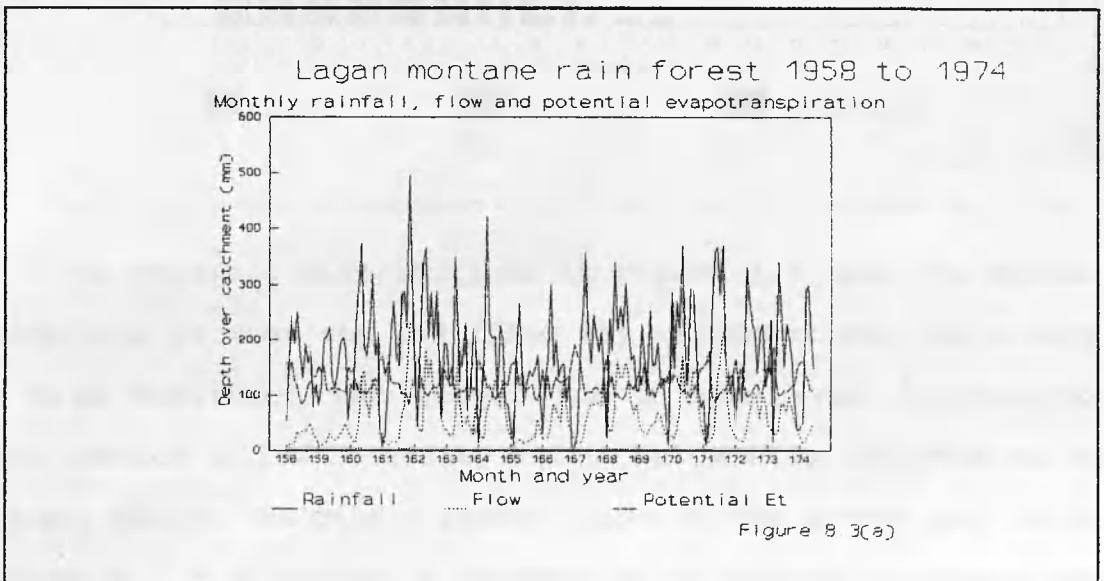
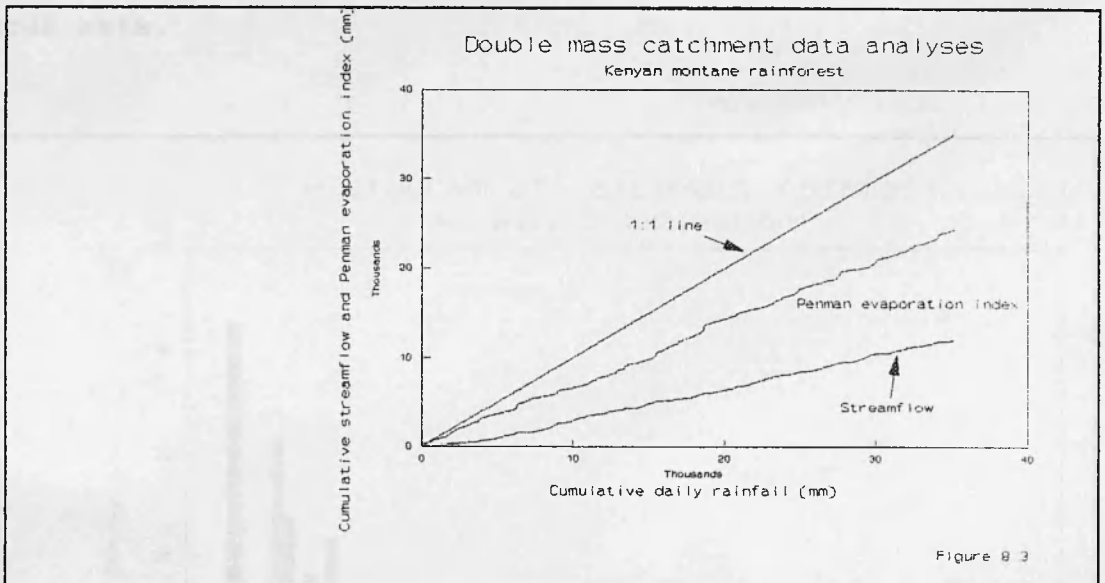
To be homogenous the sets of data should have no breaks in the overall slope of the cumulative plotted double mass curves, or continuing changes in the slope. Data sets from Plynlimon in Wales were rejected for this study due to breaks in the curves from the 1975/76 drought, which appears to have changed the upland peat soil storage until it recovered in 1981. A similar problem developed in the summer of 1984.

The double mass curves and time series plots are given for each catchment in Figures 8.1-8.3, and consist of cumulative plots of daily



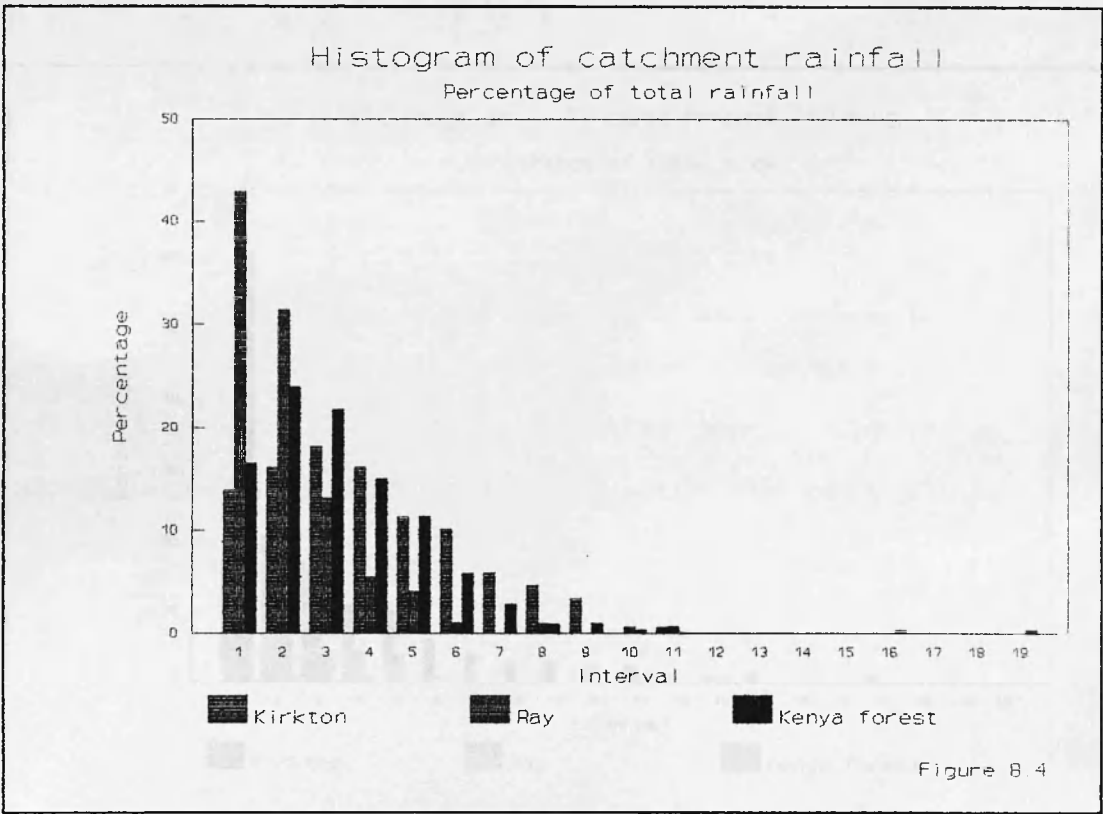
rainfall against streamflow, and Penman Open Water potential as an index of evaporative demand. They are cumulative over different lengths of time: the Kirkton has five years of data, the Ray seventeen years and Kenyan montane forest on the Lagan river sixteen years.

These homogeneity criteria for the three catchments are met for each of the given plots, and their rainfall/runoff and rainfall/evaporative potential are self consistent over each run of data. The evaporative loss index Penman E_0 is the most sensitive to dry seasons and the Ray data has a particularly marked step over the 1975/76 drought.



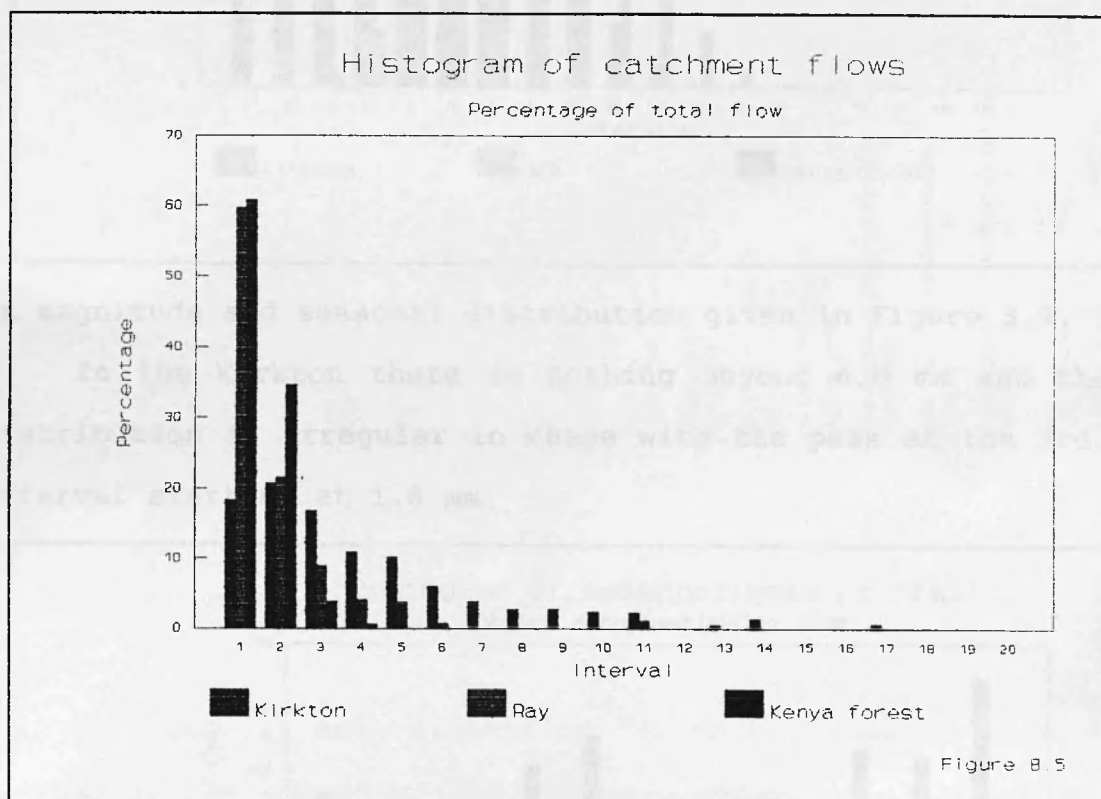
8.6 Frequency distributions of data

The three sets of data for each catchment are given for comparison as histograms in Figure 8.4-8.6 with standardised intervals for comparison: Rainfall 6.6 mm, Flow 3.6 mm, and Penman potential evaporation loss 0.6 mm. The intervals are taken from a tenth of the highest daily values from all of the three sets.

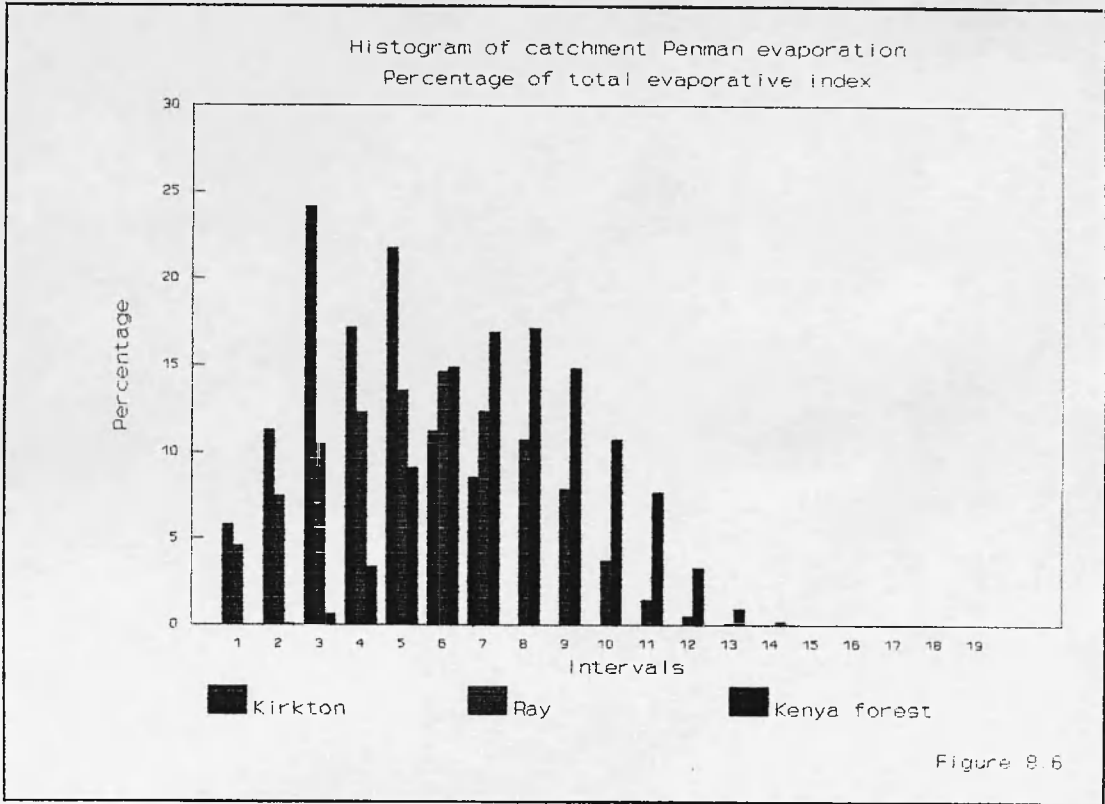


The rainfall distributions in Figure 8.4 show the marked difference between the low lying Ray catchment and the others at high altitude; the former has a more even diminishing distribution of low intensity events below 40 mm with 43% below 6.6 mm, whilst the others reach a peak on the second and third intervals. High intensity storms are in intervals sixteen and nineteen on the Kenyan catchment but only form 0.66% of the total rainfall input.

The flow distributions given in Figure 8.5 show a low flow percentage of 61% and nothing above the 4th. interval 14.4 mm, with the Ray 60% in the first interval extending to the sixth interval with a single storm in the 11th. interval. The Kirkton distribution rises to a peak of 21% in the second interval with flows tailing off to a single convective storm in the seventeenth interval of 60.1 mm.

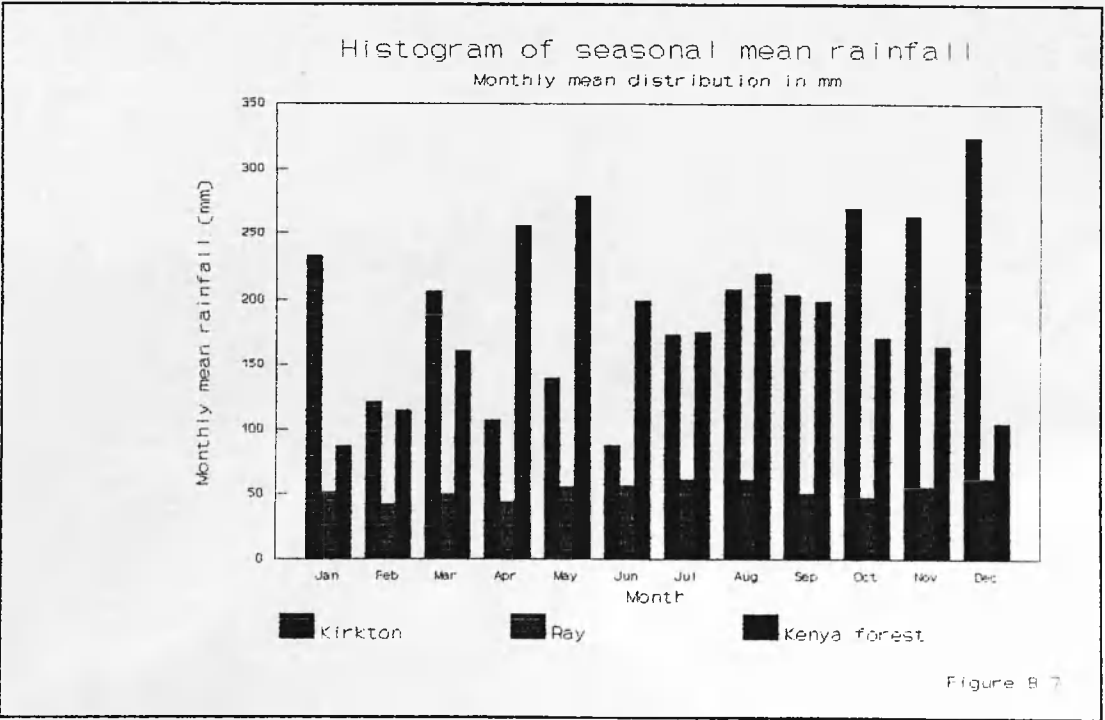


The catchments distribution of Penman open water potential evaporation given in Figure 8.6 shows an interesting similarity between the Ray and Kenyan catchments in that they are regular in shape and cover twelve intervals with their peaks lying in the sixth interval but displaced by two intervals. This is rather unexpected considering their difference in mean altitude, 2,100 m, and latitude of 60°, and the wide difference

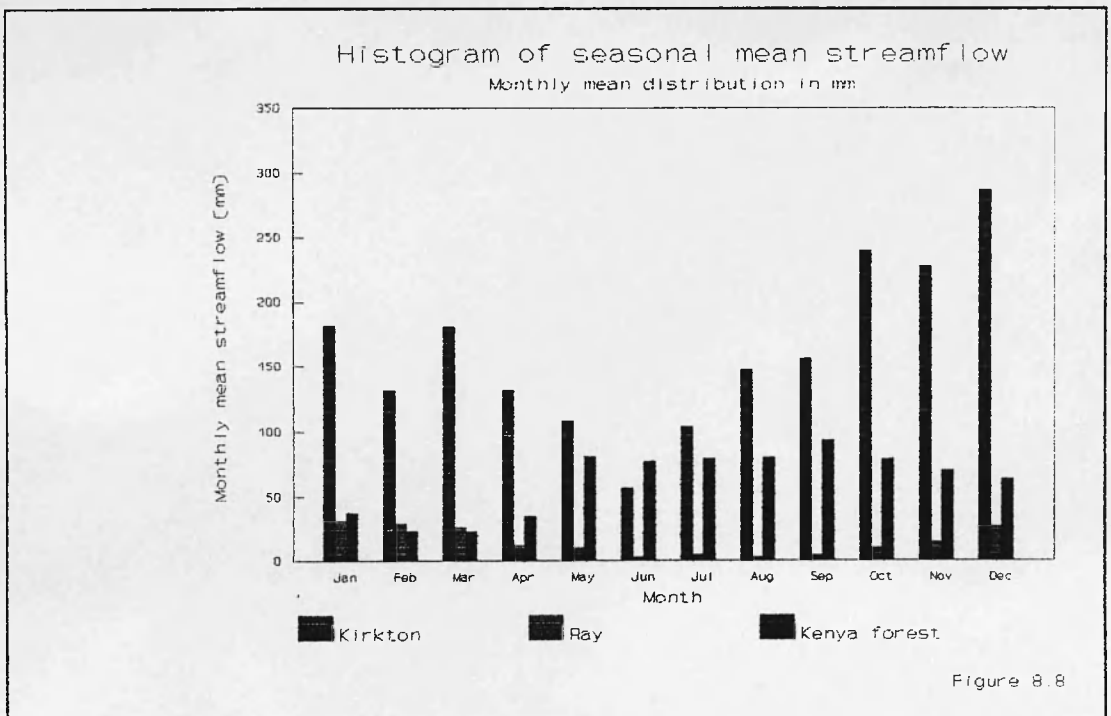


in magnitude and seasonal distribution given in Figure 8.9.

In the Kirkton there is nothing beyond 4.8 mm and the distribution is irregular in shape with the peak at the 3rd. interval starting at 1.8 mm.



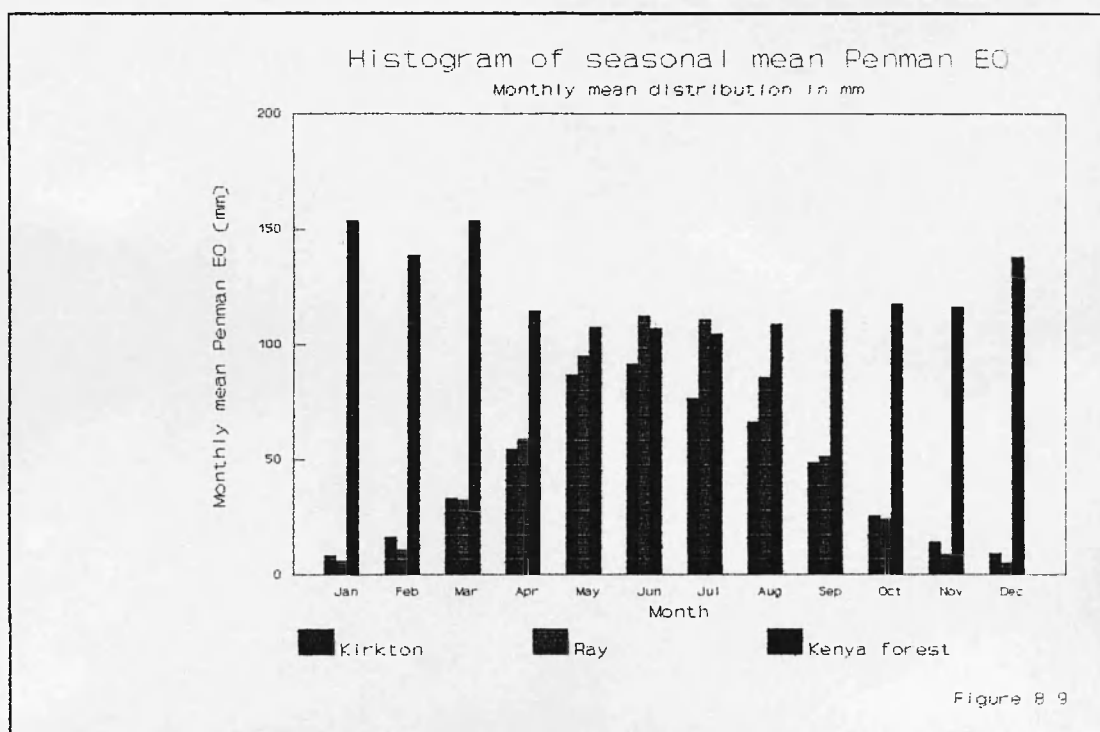
The seasonal mean rainfall given in Figure 8.7 for the Ray catchment shows little variation throughout the year being at its wettest in July and August, and driest in February and April. The Kenyan forest catchment has a bimodal seasonal distribution of rainfall with the 'long' rains in April/May and the 'short' rains in August/September. The Kirkton has its peak in December and minimum in June.



The seasonal mean streamflow for Kenya, given in Figure 8.8, shows the effect of the two rain seasons with the first peak in May and little change until the highest peak in September; the flows then decrease until the minimum in February and March. The Ray has relatively small flows with the peak in December/January and flows ceasing altogether at times in June through to August. Flows are higher than the other two catchments throughout most of the average year in the

Kirkton with peak flows in December and minimum in June which is below the May/June peak flows in Kenya.

The histogram of seasonal mean Penman E_0 is given in Figure 8.9 showing the same relative pattern for the Kirkton and the Ray catchments with a higher potential evaporative demand in the Kirkton from September through to March possibly due to higher wind speeds in the North. The potential in Kenya near the equator tends to be more evenly distributed throughout the year with the maximum demand in the long dry season in December to March.



8.7 River stage recession curves

A computer program, BASEFLOW, was written to search for flow recession data i.e decreasing runs of flow data which are taken to represent the output from groundwater storage with less than 2 mm rainfall input. It is sometimes possible to

analyze these curves to obtain estimates of the ground water storage parameters. The parameters are p_{21} and p_{22} ; due to the 'top/down' structure of the model tend, when optimized, to accumulate the errors and bias stemming from problems with 'higher' parameters in the structure; it is therefore important, if possible, to obtain an estimate of these parameters from the observed data.

Assume that the non-linear reservoir equation applies

$$\begin{aligned} Q_g &= AG_s^B \\ &= \frac{\partial G_s}{\partial t} \end{aligned} \quad (8.1)$$

and Q_g is the groundwater storage output, G_s is groundwater storage, A and B are constants shown as p_{21} and p_{22} in the model parametric structure, and the partial differential is with respect to time. It was explained at the beginning of this section that the baseflow recession is assumed to be entirely from the groundwater store, and is a monotonic decreasing function of time. It is a continuous decreasing function of time as indicated by the minus sign in equation 8.1.

This expression has the form when inverted to obtain groundwater storage as the dependent variable

$$y = \alpha \cdot x^\beta \quad \text{where } \alpha = \frac{1}{A}, \quad \beta = \frac{1}{B} \quad (8.2)$$

here y is groundwater storage at time t , α and β are the inverse constants to those in (8.1), and x is the outflow at time t given by

$$x = -\frac{\partial y}{\partial t} \quad (8.3)$$

and outflow loss rate is given by (8.2). Partial differentiation of (8.2) with respect to time gives

$$x^{2-\beta} = -\alpha \cdot \beta \cdot \frac{\partial x}{\partial t} \quad (8.4)$$

which expressed in logarithmic form is

$$(2 - \beta) \cdot \text{Log} \cdot x = \text{Log}(\alpha \cdot \beta) + \text{Log}\left(-\frac{\partial x}{\partial t}\right) \quad (8.5)$$

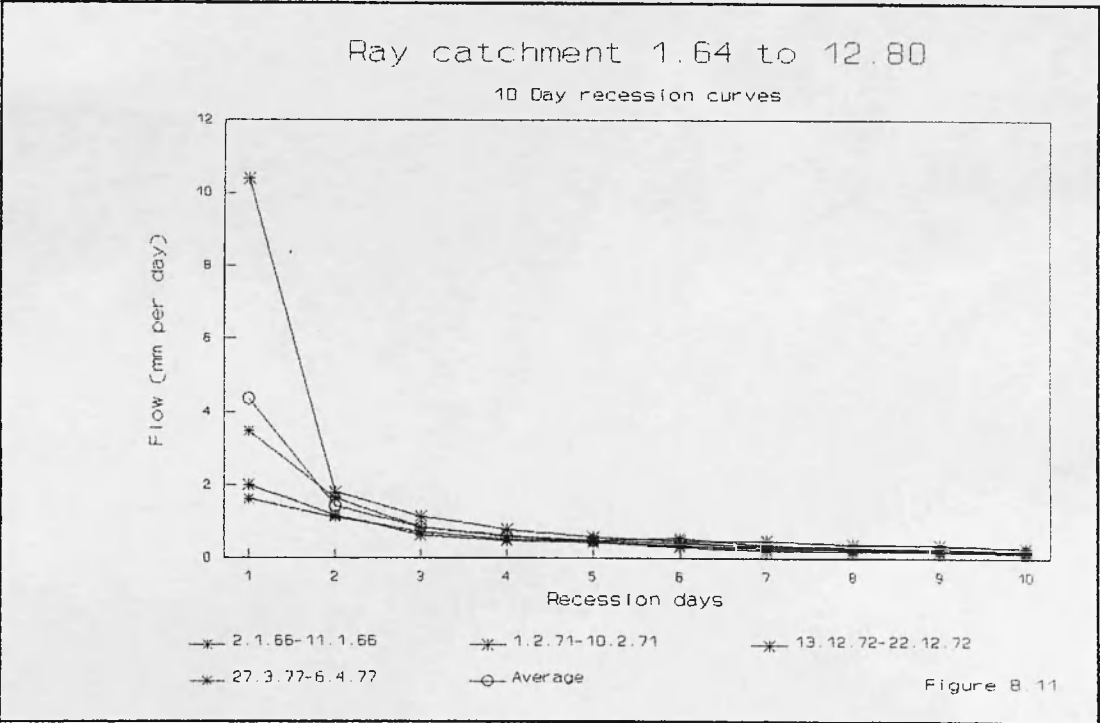
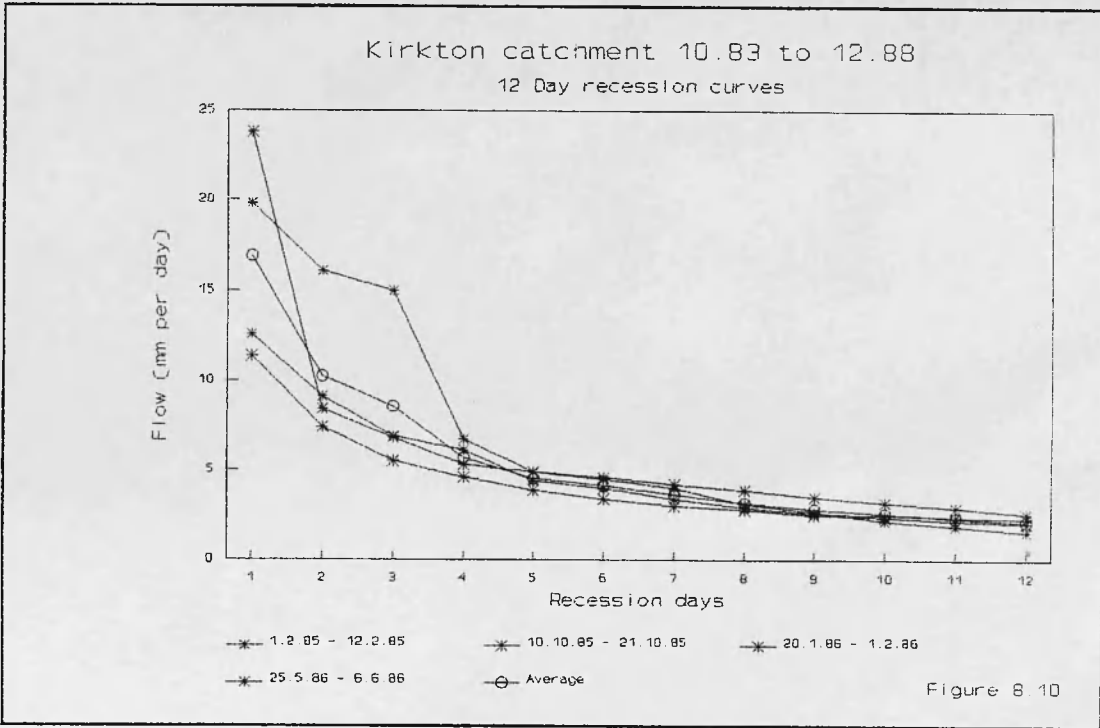
and has the linear logarithmic form

$$\text{Log} x = \frac{1}{2-\beta} \cdot \text{Log}(\alpha \cdot \beta) + \frac{1}{2-\beta} \cdot \text{Log}\left(-\frac{\partial x}{\partial t}\right) \quad (8.6)$$

From a plot of (8.6) values of constants α and β can be found, and hence p_{21} and p_{22} from (8.1). This method unfortunately requires the manual extraction of a large number of tangents to the recession curves by not very accurate geometrical methods, and so it would appear to be quicker to consider a geometrical regression technique as in the following section.

8.8 Plots of recession curves and regressions

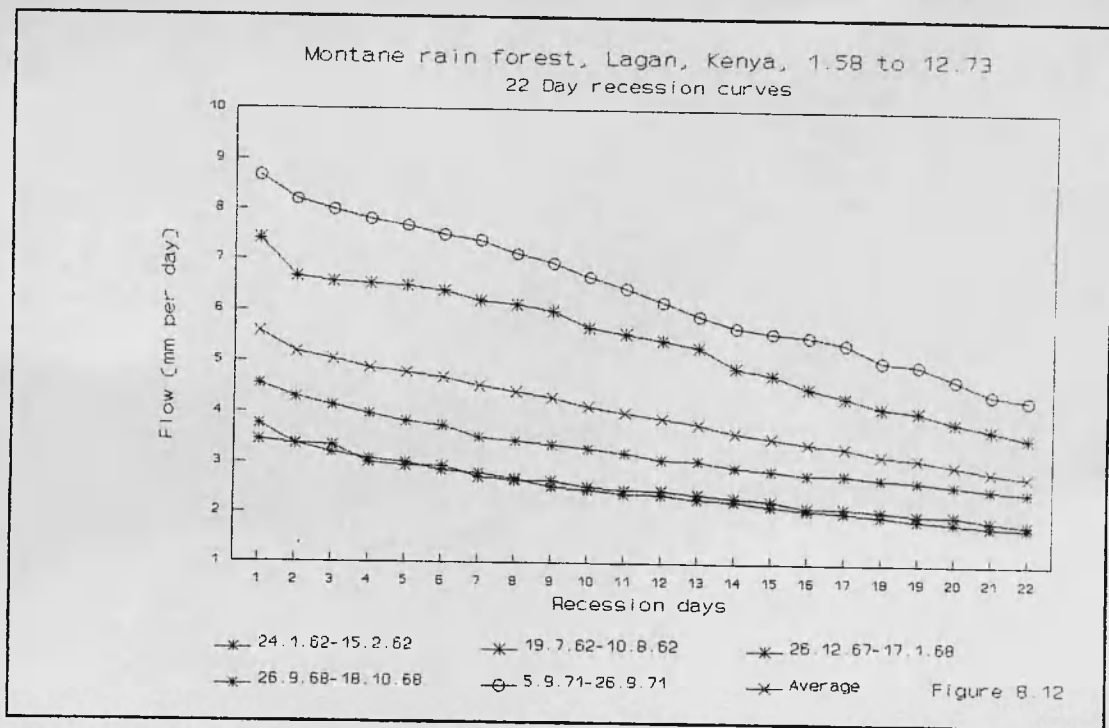
The four 12 day recession curves for the Kirkton are given in Figure 8.10 together with their average. The four curves were found by the program BASEFLOW from the flow data from October 1983 to December 1988. The average is probably biased by the data for 25.5.86 to 6.6.86 which is affected by surface runoff in that period.



The four ten day curves for the Ray from the period 1.64 to 12.80 are given in Figure 8.11 and are remarkably uniform after the second day of recession implying a rapid response to surface runoff.

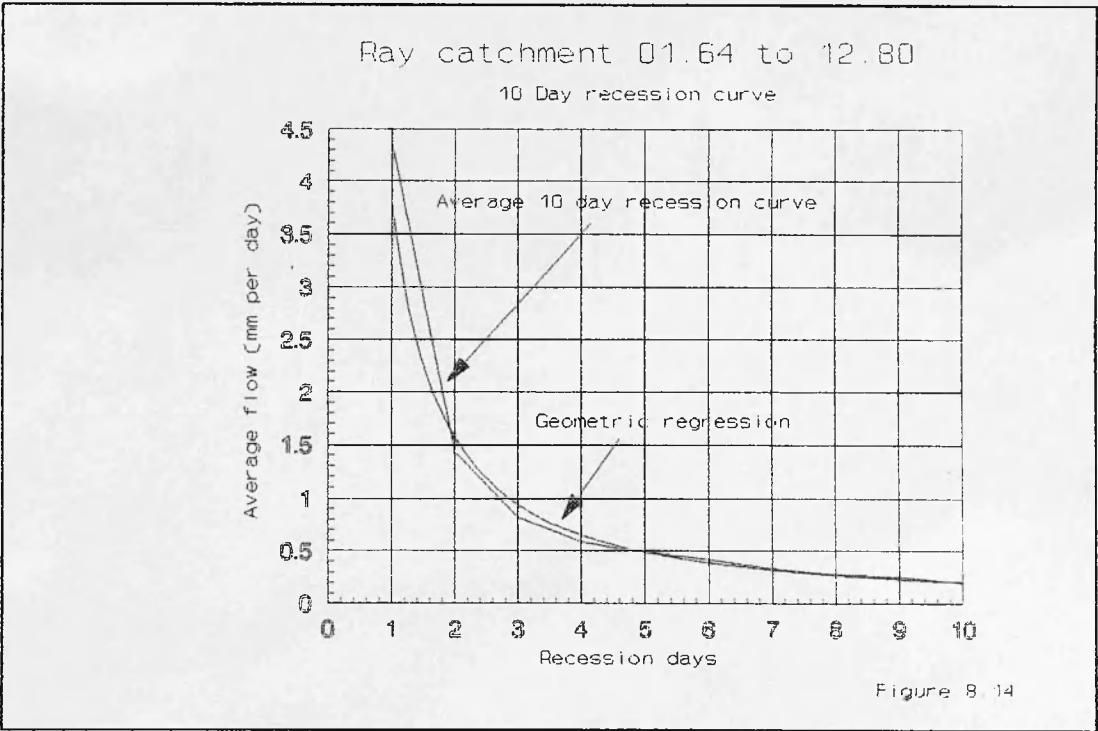
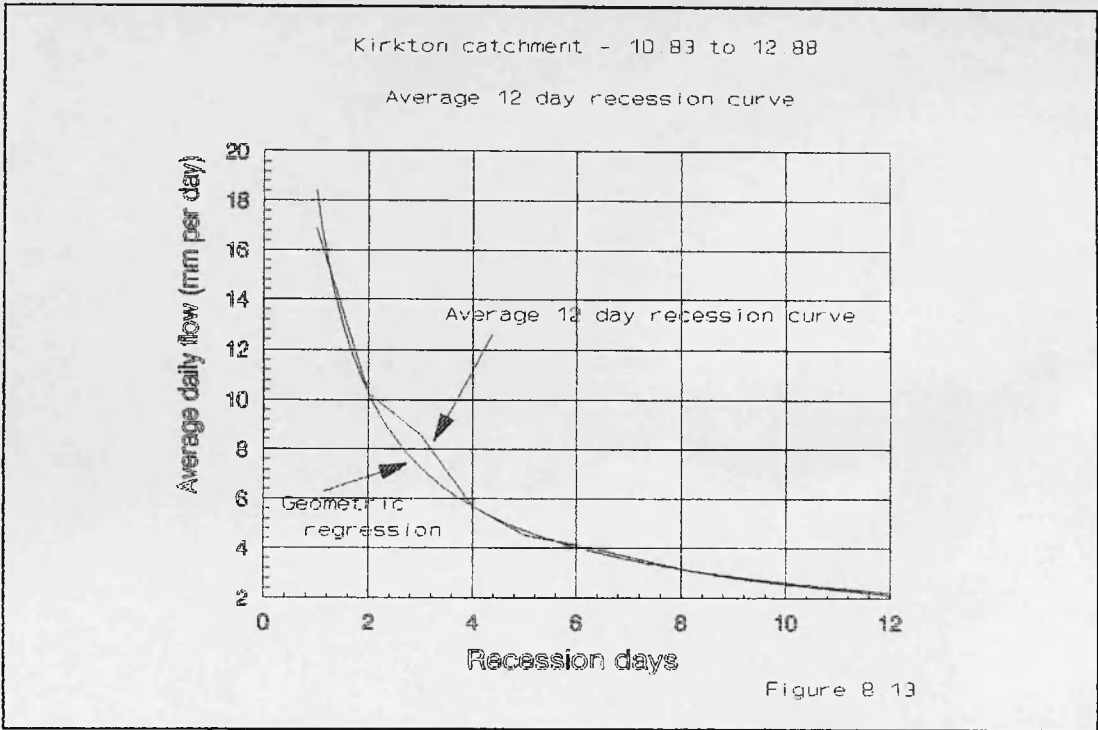
There are five 22 day recession curves for the Kenya catchment for the period 1.58 to 12.73 given in Figure 8.12. The curves show that they could possibly be overlapped to form a longer curve, and have later been combined to give a 45 day recession curve.

The three Figures 8.13-8.15 give plots of the three average recessions with geometric regressions fitted; the regression for the Kenyan catchment is poorly correlated and so

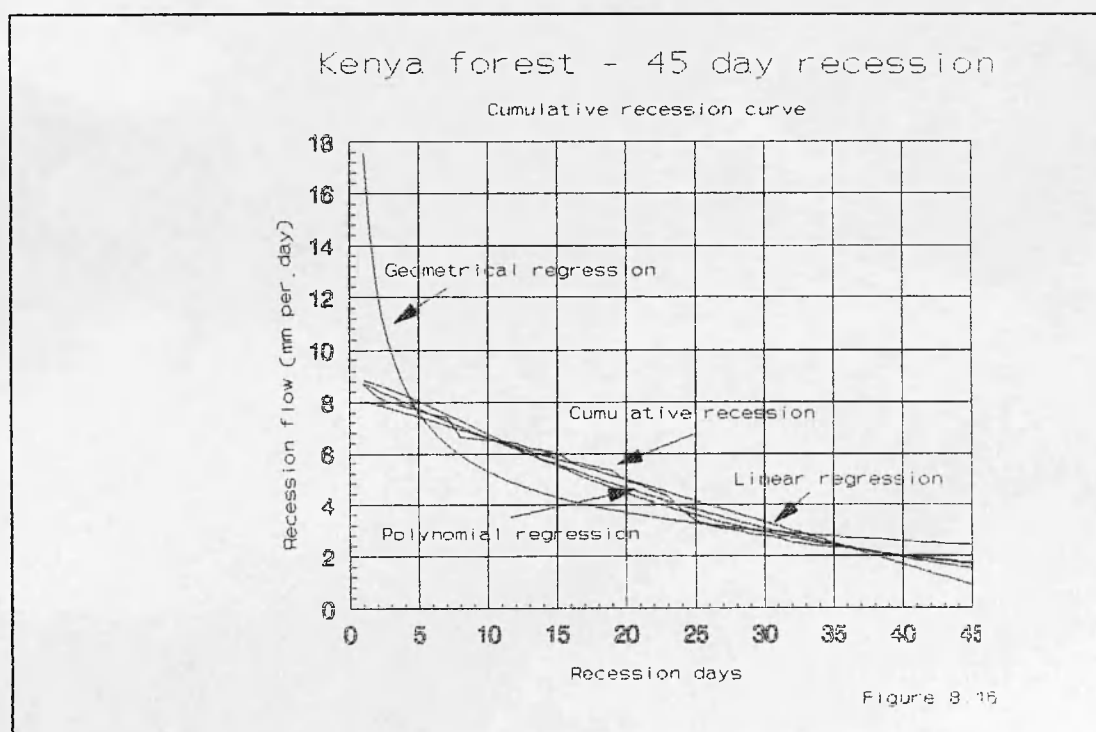
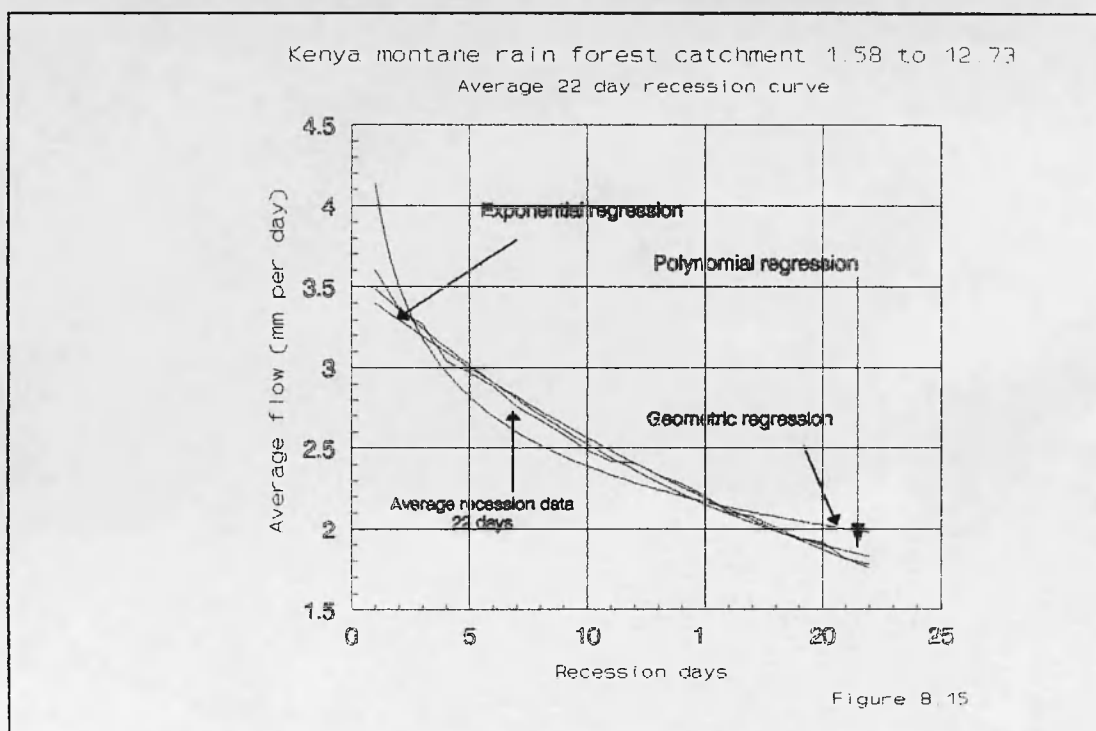


other forms of curve were fitted which gave much the same improved correlation for the 22 day recessions.

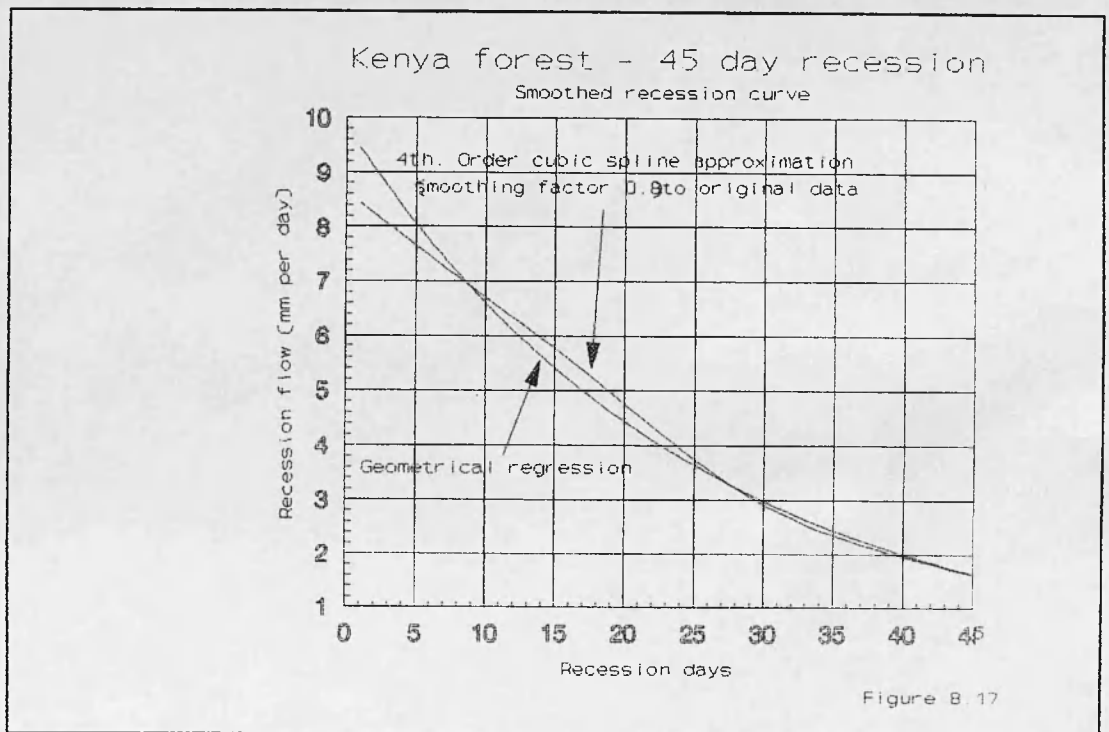
A cumulative 45 day recession curve is plotted in Figure 8.16 but again the geometrical regression is not a particularly good fit to the observed data. In Figure 8.17 the 45 day



recession curve of observed data is shown when approximated by a 4th. order cubic spline with a smoothing factor of 0.8. The program used for this approximation was taken from the GRAFTOOL package published by the 3-D Vision Corporation of America.



Yet again the geometric regression was not successful, but gave a good fit when the last 24 days of the curve were used for the regression. The results of these regression analyses are shown in Table 8.7 at the end of this Chapter. This Table 8.7 gives an idea of the wide variation between parameters with



different catchments and length of days observable in the recession.

The model is not the 'real' world and is only a simulation of it with a wide range of parameter values which will give the same degree of fit to the observed hydrograph. Therefore if information can be extracted from observed data that can be used to fix, or give an initial starting value, to a parameter then the model developed is likely to be robust in the validation period.

Table 8.1.
Kirkton Catchment monthly rainfall, streamflow, and Penman E.

YEAR	MONTH												TOTALS				
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	RAIN	R DAYS	FLOW	E	
1984	R	357.7	166.0	161.1	78.5	25.0	77.0	58.0	63.6	220.9	325.5	416.3	265.5	2215.1			
	RD	29	23	16	15	13	21	13	13	27	30	27	29		256		
	Q	181.7	165.6	145.5	171.5	34.0	24.2	18.4	15.2	101.9	271.9	378.9	262.3			1771.1	
	E	8.4	18.9	30.7	64.7	104.3	101.7	113.1	86.5	52.4	30.9	13.5	9.9				635.0
	R ² Q	176.0	.4	15.6	-93.0	-9.0	52.8	39.6	48.4	119.0	53.6	37.4	3.2	R-Q	444.0		
1985	R	81.5	78.4	107.2	166.9	112.4	86.6	261.7	374.0	272.2	182.5	202.0	359.9	2285.3			
	RD	15	15	25	23	19	22	26	30	25	12	22	27		261		
	Q	73.7	113.9	86.4	151.3	84.8	48.7	176.7	302.1	227.6	230.6	111.8	351.5			1959.1	
	E	2.8	12.2	23.8	52.3	80.5	82.3	65.0	55.1	36.2	26.7	5.7	3.1				445.7
	R ² Q	7.8	-35.5	20.8	15.6	27.6	37.9	85.0	71.9	44.6	-48.1	90.2	8.4	R-Q	326.2		
1986	R	326.7	18.9	338.2	109.2	378.2	88.9	109.9	182.2	51.4	281.8	414.5	489.5	2789.4			
	RD	25	19	26	23	29	15	24	20	16	28	29	30		284		
	Q	216.7	36.9	301.1	103.9	327.8	74.9	51.7	138.1	26.2	165.7	392.3	407.9			2243.2	
	E	13.3	4.8	40.0	56.4	79.1	113.4	68.2	62.0	53.9	36.9	25.8	4.4				558.2
	R ² Q	110.0	-18.0	37.1	5.3	50.4	14.0	58.2	44.1	25.2	116.1	22.2	81.6	R-Q	546.2		
1987	R	88.6	120.1	181.5	78.3	83.9	148.9	98.9	164.2	250.0	250.9	149.9	283.9	1899.1			
	RD	18	19	22	17	15	19	21	25	22	24	15	18		235		
	Q	142.0	117.6	161.8	104.1	41.8	98.7	60.8	95.6	215.9	220.8	143.1	207.8			1610.0	
	E	4.0	16.2	31.0	51.0	90.2	67.8	80.0	71.2	52.8	13.4	9.8	4.1				491.5
	R ² Q	-53.4	2.5	19.7	-25.8	42.1	50.2	38.1	68.6	34.1	30.1	6.8	76.1	R-Q	289.1		
1988	R	310.0	223.1	242.5	100.9	96.2	35.8	337.1	256.4	224.9	309.5	133.8	222.6	2492.8			
	RD	27	24	26	18	17	14	31	25	29	23	13	26		273		
	Q	294.2	223.4	210.1	126.5	54.9	36.8	211.7	185.0	208.5	307.9	111.0	201.1			2171.1	
	E	11.8	29.9	40.3	47.1	77.2	91.2	55.2	55.4	47.4	20.3	15.8	24.5				516.1
	R ² Q	15.8	-.3	32.4	-25.6	41.3	-1.0	125.4	71.4	16.4	1.6	22.8	21.5	R-Q	321.7		
MEAN	R	232.9	121.3	206.1	106.8	139.1	87.4	173.1	208.1	203.9	270.0	263.3	324.3	2336.3			
	RD	22	20	23	19	18	18	23	22	23	23	21	26		261		
	Q	181.7	131.5	181.0	131.5	108.7	56.7	103.9	147.2	156.0	239.4	227.4	286.1			1950.9	
	E	8.1	16.4	33.2	54.3	86.3	91.3	76.3	66.0	48.5	25.6	14.1	9.2				529.3
	R ² Q	51.2	-10.2	25.1	-24.7	30.5	30.8	69.3	80.9	47.9	30.7	35.9	38.2	R-Q	385.4		
S.D.	R	136.1	78.6	88.3	36.3	137.6	40.5	119.9	115.5	87.7	56.6	141.1	104.9	331.0			
	RD	6	3	4	3	6	3	6	6	5	7	7	4		18		
	Q	82.3	69.0	80.4	29.7	124.0	30.0	84.9	106.8	88.4	53.9	145.1	90.9			265.8	
	E	4.6	9.2	7.0	6.7	11.3	17.5	22.4	13.2	7.3	9.1	7.6	9.0				71.8
	R ² Q	91.0	16.4	9.2	42.4	23.5	23.5	36.6	13.5	41.1	61.0	32.2	37.8	SD R-Q	107.4		

R is monthly total of rainfall
RD is number of days on which there was rain
Q is monthly total of streamflow
E is the monthly total of the Penman open water evaporation index
R²Q is the total monthly change in soil moisture and groundwater storage, and actual evapotranspiration loss
S.D. is the standard deviation of the above variables for the period 1.84 to 12.88

Table 8.2

Kirkton catchment daily frequency distributions 1984 to 1988.

INTERVAL	RAIN				FLOW				EVAP			
	Interval 6.6 mm				Interval 3.6 mm				Interval 0.6 mm			
	mm	%	Daily	%	mm	%	Daily	%	mm	%	Daily	%
1	1617.8	13.85	787	60.12	1795.4	18.41	1003	54.90	153.2	5.79	514	28.13
2	1883.3	16.12	198	15.13	2027.1	20.78	402	22.00	299.4	11.31	336	18.39
3	2117.0	18.12	129	9.85	1640.6	16.82	188	10.29	639.1	24.15	395	21.62
4	1881.7	16.11	82	6.26	1079.0	11.06	89	4.87	455.3	17.20	214	11.71
5	1318.3	11.29	45	3.44	996.0	10.21	62	3.39	575.8	21.76	216	11.82
6	1176.0	10.07	33	2.52	589.3	6.04	30	1.64	297.2	11.23	91	4.98
7	683.8	5.85	16	1.22	375.3	3.85	16	.88	226.5	8.56	61	3.34
8	544.5	4.66	11	.84	266.8	2.74	10	.55	.0	.00	0	.00
9	390.2	3.34	7	.53	271.8	2.79	9	.49	.0	.00	0	.00
10	.0	.00	0	.00	239.8	2.46	7	.38	.0	.00	0	.00
11	69.1	.59	1	.08	224.2	2.30	6	.33	.0	.00	0	.00
12	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
13	.0	.00	0	.00	46.1	.47	1	.05	.0	.00	0	.00
14	.0	.00	0	.00	143.0	1.47	3	.16	.0	.00	0	.00
15	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
16	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
17	.0	.00	0	.00	60.1	.62	1	.05	.0	.00	0	.00
18	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
19	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
20	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
TOTALS	11681.7		1309		9754.5		1827		2646.5		1827	

Table 8.3
Ray Catchment Monthly rainfall, streamflow and Penman E_o

YEAR		MONTH												TOTALS			
		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	RAIN	RDAYS	FLOW	E _o
1964	R	15.8	20.4	86.3	58.4	42.1	74.4	91.6	18.2	18.5	20.6	21.6	42.0	509.8			
	RD	13	12	17	18	16	14	16	13	6	16	11	18		170		
	Q	9.7	8.9	62.6	19.0	2.2	3.6	10.9	.0	.0	.0	.0	.0			116.7	
	E	3.9	9.9	21.3	61.0	110.9	104.3	120.2	94.4	63.1	19.8	7.1	5.6				621.5
	R ² Q	6.1	11.5	23.8	39.4	39.9	70.8	80.7	18.2	18.5	20.6	21.6	42.0	R-Q	393.0		
1965	R	56.2	11.0	48.0	50.3	55.6	64.1	105.2	54.4	82.7	10.3	54.9	99.6	692.3			
	RD	20	14	13	22	18	15	20	16	19	7	21	22		207		
	Q	6.1	1.1	16.3	4.4	4.6	.6	9.0	1.6	11.7	3.1	14.6	73.5			146.7	
	E	5.9	13.2	35.2	55.4	95.0	116.7	100.0	91.3	50.9	24.7	7.0	3.4				598.7
	R-Q	50.0	9.9	31.8	46.0	51.0	63.5	96.2	52.7	70.9	7.2	40.3	26.1	R-Q	545.6		
1966	R	32.9	77.8	15.1	84.4	66.0	65.6	64.4	101.6	44.2	119.3	41.6	76.1	789.0			
	RD	16	18	12	20	12	17	16	16	9	18	23	24		201		
	Q	28.8	59.7	5.2	45.7	22.2	.9	.1	6.7	1.1	64.0	14.9	53.3			302.6	
	E	5.1	16.1	35.6	51.6	107.1	123.2	107.7	91.3	55.4	20.2	8.0	3.8				625.1
	R ² Q	4.1	18.1	9.9	38.7	43.9	64.7	64.3	94.9	43.1	55.3	26.6	22.8	R-Q	486.5		
1967	R	31.5	60.6	29.7	43.4	104.6	32.5	88.0	33.2	55.0	115.9	33.3	57.9	685.6			
	RD	14	13	12	15	25	7	14	12	17	24	16	22		191		
	Q	25.1	33.9	19.4	5.9	22.6	1.1	1.8	.0	.0	17.4	20.9	33.2			181.3	
	E	4.2	13.6	43.4	56.9	90.5	116.0	125.3	89.8	44.5	25.0	3.9	3.3				616.3
	R ² Q	6.4	26.6	10.3	37.5	82.1	31.4	86.3	33.1	55.0	98.5	12.4	24.6	R-Q	504.2		
1968	R	56.2	16.5	25.7	50.7	64.0	65.2	121.5	71.9	114.8	53.9	57.5	63.9	761.7			
	RD	18	11	14	14	20	17	12	16	20	16	17	13		188		
	Q	49.2	9.7	2.0	3.3	14.8	1.0	44.7	3.1	31.9	23.1	35.4	44.7			262.9	
	E	5.3	7.5	42.0	63.6	82.4	104.7	94.6	65.3	50.8	24.0	7.0	2.3				549.4
	R ² Q	7.0	6.7	23.7	47.4	49.2	64.1	76.8	68.8	82.9	30.8	22.1	19.3	R-Q	498.9		
1969	R	70.7	36.7	48.5	33.3	104.7	24.7	43.7	69.7	12.0	3.5	52.0	60.7	560.0			
	RD	23	15	13	16	.22	10	7	18	8	5	14	19		170		
	Q	64.1	32.0	33.2	4.8	25.3	4.0	.2	2.2	.0	.0	.2	7.8			173.8	
	E	4.4	7.1	22.3	64.3	91.6	144.2	129.4	81.4	47.1	25.4	10.5	3.4				631.1
	R ² Q	6.6	4.6	15.3	28.4	79.4	20.8	43.6	67.5	12.0	3.5	51.7	52.9	R-Q	386.3		
1970	R	60.9	51.5	42.4	81.4	19.2	32.7	50.9	66.8	31.8	19.6	142.7	37.4	637.5			
	RD	25	17	18	23	8	10	15	12	11	9	23	14		185		
	Q	28.1	34.9	27.9	40.5	2.0	.0	.0	.2	.0	.0	16.4	20.0			170.0	
	E	4.2	9.2	31.4	52.4	106.0	135.7	108.4	84.9	53.9	25.8	9.0	5.8				626.7
	R ² Q	32.8	16.6	14.5	40.9	17.3	32.6	50.9	66.6	31.8	19.6	126.3	17.4	R-Q	467.4		
1971	R	95.8	16.9	42.8	41.1	28.7	97.7	43.4	98.1	16.9	75.5	76.2	22.8	655.8			
	RD	20	11	21	8	14	15	10	20	5	13	15	5		157		
	Q	61.4	22.8	21.0	8.8	.3	9.0	.2	3.4	.0	8.5	30.6	11.2			177.3	
	E	4.1	9.9	30.9	54.5	102.2	91.2	127.8	79.9	52.8	21.5	6.4	5.6				586.7
	R ² Q	34.4	-5.9	21.8	32.3	28.3	88.6	43.2	94.7	16.9	67.0	45.6	11.6	R-Q	478.4		
1972	R	62.7	46.5	51.5	45.8	52.9	36.4	62.8	33.9	31.4	19.2	63.5	69.4	576.1			
	RD	22	20	17	20	22	17	11	6	8	9	18	17		187		
	Q	26.7	34.8	24.4	8.7	4.7	.1	.1	1.3	.1	.0	2.1	27.6			130.6	
	E	6.2	6.7	36.4	60.1	92.5	99.2	100.9	78.5	45.1	22.8	6.5	6.2				561.0
	R ² Q	36.0	11.7	27.2	37.0	48.2	36.3	62.7	32.6	31.3	19.2	61.4	41.9	R-Q	445.5		
1973	R	23.0	18.4	13.9	43.6	51.8	70.7	73.1	34.9	33.0	24.7	36.5	40.3	464.0			
	RD	13	18	8	16	19	8	11	10	13	13	13	16		158		
	Q	8.7	6.2	2.3	1.2	3.5	1.6	3.9	.2	.0	.0	.0	2.7			30.3	
	E	4.4	11.1	35.2	61.7	97.2	122.6	100.3	108.3	51.3	20.6	9.5	6.2				628.4
	R ² Q	14.3	12.1	11.6	42.4	48.3	69.1	69.2	34.7	33.0	24.7	36.5	37.7	R-Q	433.7		

1974	R	64.2	72.6	44.0	7.0	31.5	89.8	42.8	92.0	118.7	76.3	114.7	30.5	784.1	
	RD	20	19	14	4	11	11	16	14	21	17	24	16	187	
	Q	29.4	41.7	22.3	.3	.1	1.3	.3	.6	17.8	39.2	79.8	13.6	246.4	
	E	6.6	13.3	26.6	57.4	104.1	112.3	100.0	76.7	45.1	17.8	13.1	9.4	582.4	
	R ² Q	34.8	31.0	21.7	6.7	31.4	88.5	42.5	91.4	100.9	37.1	34.8	16.9	R-Q 537.7	
1975	R	68.0	30.1	87.3	46.0	40.1	10.1	58.1	25.7	78.4	10.8	39.3	23.2	517.4	
	RD	21	8	27	21	12	4	16	10	15	10	16	12	172	
	Q	43.7	23.7	57.9	21.4	3.8	.0	.5	.0	.2	.0	.1	.3	151.7	
	E	9.2	10.9	24.7	56.9	88.5	138.5	131.0	108.6	53.9	21.5	6.2	3.3	653.2	
	R ² Q	24.3	6.4	29.5	24.6	36.2	10.1	57.6	25.7	78.2	10.8	39.3	22.9	R-Q 365.7	
1976	R	18.2	20.2	16.3	9.0	28.1	16.0	17.8	16.4	100.7	92.3	55.5	87.2	477.5	
	RD	13	14	10	6	17	7	10	3	17	26	18	20	161	
	Q	.6	.6	.1	.0	.0	.0	.0	.0	.3	2.6	5.7	48.4	58.3	
	E	7.4	10.9	36.3	65.9	105.4	141.7	134.9	96.9	45.6	25.1	5.0	3.2	678.3	
	R ² Q	17.6	19.6	16.1	9.0	28.1	16.0	17.8	16.4	100.4	89.7	49.8	38.8	R-Q 419.2	
1977	R	74.5	107.9	59.4	37.4	35.5	99.7	4.9	130.2	20.1	32.5	42.8	63.8	708.7	
	RD	26	21	23	15	12	17	9	17	12	15	13	18	198	
	Q	71.3	87.5	26.2	4.1	9.0	8.3	.1	3.9	1.8	.1	2.9	30.7	246.1	
	E	8.5	18.3	38.7	56.5	80.4	72.0	89.8	61.8	51.3	29.4	18.3	6.0	531.0	
	R ² Q	3.2	20.4	33.2	33.2	26.5	91.4	4.7	126.3	18.2	32.5	39.9	33.1	R-Q 462.6	
1978	R	63.2	42.2	60.4	42.9	49.7	38.9	64.9	32.0	30.3	2.0	23.9	93.1	543.5	
	RD	22	18	22	13	11	13	13	15	12	6	9	22	176	
	Q	40.2	29.4	28.2	8.8	34.8	.0	.1	.6	.0	.0	.0	9.1	151.1	
	E	8.7	13.9	37.2	64.9	103.8	113.4	106.9	88.7	69.3	31.2	9.4	6.0	653.5	
	R ² Q	23.0	12.8	32.2	34.2	14.9	38.9	64.8	31.3	30.3	2.0	23.8	84.0	R-Q 392.4	
1979	R	51.7	41.4	110.9	45.2	104.4	48.3	22.7	70.3	17.2	55.0	37.0	142.3	746.4	
	RD	17	12	25	18	16	9	6	19	8	12	19	23	184	
	Q	24.5	42.8	67.0	18.6	29.0	8.7	.0	.3	.0	.3	1.6	56.5	249.2	
	E	5.1	10.4	34.3	50.7	53.3	85.9	113.2	78.8	39.5	38.1	16.6	4.4	530.3	
	R ² Q	27.2	-1.3	43.9	26.6	75.4	39.5	22.7	70.0	17.2	54.8	35.4	85.8	R-Q 497.2	
1980	R	27.5	39.6	63.2	14.8	55.3	87.7	72.2	78.7	47.7	71.1	46.1	32.2	636.3	
	RD	10	13	17	7	8	20	20	14	8	17	13	15	162	
	Q	13.6	20.4	26.6	11.3	.6	2.7	3.3	13.8	2.2	12.7	21.6	18.3	147.1	
	E	6.2	4.6	22.7	62.4	103.2	89.9	88.3	74.9	53.7	19.0	3.8	1.5	530.3	
	R ² Q	13.9	19.2	36.7	3.6	54.7	85.0	68.8	64.9	45.5	58.4	24.6	13.9	R-Q 489.1	
MEAN	R	51.3	41.8	49.7	43.2	55.0	56.1	60.5	60.5	50.2	47.2	55.2	61.3	632.1	
	RD	18	14	16	15	15	12	13	13	12	13	16	17	179	
	Q	31.2	28.8	26.0	12.2	10.6	2.5	4.4	2.2	3.9	10.1	14.5	26.5	173.1	
	E	5.8	11.0	32.6	58.6	94.9	112.4	110.5	85.4	51.4	24.2	8.7	4.7	600.2	
	R ² Q	20.1	12.9	23.7	31.1	44.4	53.6	56.0	58.2	46.3	37.2	40.7	34.8	R-Q 459.0	
S.D.	R	22.6	26.2	26.9	20.8	27.0	28.6	30.7	33.1	35.6	38.6	31.3	31.7	108.1	
	RD	4	3	5	5	5	4	4	4	5	5	4	4	15	
	Q	21.1	22.1	20.2	13.3	11.7	3.2	10.9	3.5	8.7	17.7	20.4	22.1	71.7	
	E	1.7	3.5	6.9	4.8	13.9	20.8	15.0	12.9	7.2	5.0	4.1	1.9	46.6	
	R ² Q	14.1	9.4	10.0	13.3	20.2	27.1	24.8	31.5	29.9	29.1	25.4	22.1	SD R-Q 53.7	

R is monthly total of rainfall
RD is the number of days on which there was rain
Q is monthly total of streamflow
E is the monthly total of the Penman open water index
R²Q is the total monthly change in soil moisture and groundwater storage, and actual evapotranspiration loss
S.D. is the standard deviation of the above variables for the period 1.64 to 12.88

Table 8.4

Ray catchment daily frequency distributions 1.64 to 12.80

INTERVAL	RAIN				FLOW				EVAP			
	Interval 6.6 mm				Interval 3.6 mm				Interval 0.6 mm			
	mm	%	Daily	%	mm	%	Daily	%	mm	%	Daily	%
1	4608.6	42.89	2544	83.30	1755.3	59.66	6032	97.13	472.9	4.63	2219	35.73
2	3368.0	31.34	376	12.31	633.7	21.54	129	2.08	763.9	7.49	873	14.06
3	1401.4	13.04	86	2.82	262.4	8.92	30	.48	1067.0	10.46	720	11.59
4	590.9	5.50	26	.85	121.6	4.13	10	.16	1257.2	12.32	604	9.73
5	437.0	4.07	15	.49	110.8	3.76	7	.11	1384.3	13.57	510	8.21
6	105.1	.98	3	.10	21.6	.73	1	.02	1498.0	14.68	453	7.29
7	.0	.00	0	.00	.0	.00	0	.00	1264.5	12.39	327	5.27
8	98.9	.92	2	.07	.0	.00	0	.00	1098.4	10.76	245	3.95
9	.0	.00	0	.00	.0	.00	0	.00	803.4	7.87	158	2.54
10	63.0	.59	1	.03	.0	.00	0	.00	385.1	3.77	68	1.10
11	72.5	.67	1	.03	36.8	1.25	1	.02	147.4	1.44	24	.39
12	.0	.00	0	.00	.0	.00	0	.00	54.6	.53	8	.13
13	.0	.00	0	.00	.0	.00	0	.00	7.2	.07	1	.02
14	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
15	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
16	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
17	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
18	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
19	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
20	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
TOTALS	10745.5		3054		2942.1		6210		10203.9		6210	

Table 8.5
Kenyan Montane Rainforest monthly rainfall, streamflow and Penman E_o

YEAR		MONTH												TOTALS			
		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	RAIN	R DAYS	FLOW	E _o
1958	R	52.3	171.7	244.1	192.4	251.1	182.6	130.2	191.5	157.7	180.9	75.8	118.2	1948.6			
	RD	17	12	11	19	27	25	20	27	21	21	15	17	232			
	Q	9.6	10.7	12.6	9.3	31.8	29.1	35.4	32.3	41.6	46.7	28.7	19.2			307.1	
	E	158.9	139.7	159.5	126.6	118.0	92.9	83.5	94.4	116.5	125.3	130.2	134.2				1479.6
	R ² Q	42.7	161.0	231.5	183.0	219.3	153.5	94.8	159.2	116.1	134.3	47.1	99.1	R-Q	1641.5		
1959	R	82.3	112.3	201.9	205.7	222.1	131.4	111.0	140.4	193.5	204.1	182.0	53.9	1840.5			
	RD	9	12	17	24	21	22	16	24	24	26	22	8	225			
	Q	13.1	8.8	11.4	13.5	29.2	22.8	18.2	18.2	29.4	36.6	50.1	33.3			284.5	
	E	152.2	140.7	153.2	110.1	106.1	112.9	102.7	111.0	109.9	106.6	105.9	149.0				1460.2
	R ² Q	69.2	103.6	190.5	192.2	193.0	108.6	92.8	122.2	164.1	167.5	131.9	20.6	R-Q	1556.0		
1960	R	118.0	109.7	229.9	303.9	375.5	190.1	170.1	227.2	318.8	172.1	215.3	39.1	2469.7			
	RD	18	14	29	26	29	22	21	24	28	23	23	6	263			
	Q	23.0	14.4	17.7	47.0	132.3	145.4	88.9	57.4	128.7	120.8	95.5	57.5			928.6	
	E	142.2	144.9	124.6	105.8	103.5	106.4	101.6	112.0	113.9	129.2	122.1	159.3				1465.6
	R ² Q	95.0	95.3	212.2	256.9	243.2	44.7	81.2	169.8	190.2	51.3	119.8	-18.4	R-Q	1541.1		
1961	R	11.5	28.0	129.8	213.5	269.8	165.0	152.1	281.9	287.0	221.6	494.8	284.1	2539.1			
	RD	4	8	16	19	26	26	23	28	28	28	29	21	256			
	Q	26.1	12.8	10.7	12.0	20.0	14.7	18.9	62.5	76.6	130.2	219.9	245.5			849.8	
	E	170.3	135.7	160.9	124.9	122.0	120.7	120.3	103.4	103.1	100.9	83.9	120.6				1466.7
	R ² Q	-14.6	15.3	119.1	201.5	249.7	150.3	133.3	219.4	210.4	91.4	274.9	38.7	R-Q	1689.4		
1962	R	147.4	42.3	181.7	329.9	364.6	203.3	285.0	199.6	287.3	196.6	57.0	193.6	2488.3			
	RD	18	6	17	26	29	24	23	25	26	27	15	21	257			
	Q	133.7	52.5	34.2	37.7	183.5	131.4	164.3	103.7	107.7	98.2	58.4	38.7			1144.0	
	E	141.4	168.4	151.0	113.6	102.9	98.9	107.1	115.9	111.8	120.4	133.2	146.3				1511.1
	R ² Q	13.7	-10.2	147.5	292.2	181.2	71.9	120.7	95.9	179.6	98.4	-1.4	154.9	R-Q	1344.3		
1963	R	190.4	115.8	118.5	348.7	236.9	125.0	171.5	231.4	71.1	116.3	221.0	186.9	2133.4			
	RD	19	14	14	27	23	22	24	22	11	24	23	18	241			
	Q	37.3	28.2	25.5	44.4	179.4	78.9	44.0	46.9	60.4	37.2	29.1	83.4			694.7	
	E	127.6	129.6	153.5	94.8	105.0	108.1	107.0	109.8	128.4	131.9	106.0	117.9				1419.6
	R ² Q	153.1	87.6	93.0	304.3	57.5	46.2	127.4	184.4	10.7	79.1	191.9	103.4	R-Q	1438.6		
1964	R	12.6	93.2	174.9	421.8	213.9	201.4	206.7	96.3	226.8	201.1	85.8	85.1	2019.6			
	RD	7	22	18	22	26	29	31	27	27	27	18	20	274			
	Q	45.8	24.3	20.6	68.0	87.3	97.5	127.2	108.5	80.2	129.9	73.8	37.3			900.6	
	E	167.0	142.5	146.3	112.8	117.3	102.2	93.4	96.1	102.3	118.3	123.8	116.8				1438.8
	R ² Q	-33.1	68.8	154.2	353.8	126.6	103.9	79.5	-12.2	146.6	71.2	11.9	47.7	R-Q	1119.0		
1965	R	38.7	8.8	165.5	266.0	157.7	79.3	121.0	144.0	134.3	147.1	171.8	106.8	1541.0			
	RD	8	6	13	23	22	18	18	21	23	22	21	19	214			
	Q	23.2	12.6	13.7	17.6	20.0	13.3	13.8	18.4	19.9	22.1	57.4	36.2			268.2	
	E	154.3	158.7	167.7	118.1	119.3	117.6	107.9	114.4	118.9	94.2	90.2	105.6				1466.9
	R ² Q	15.5	-3.8	151.8	248.4	137.7	66.0	107.2	125.7	114.4	124.9	114.4	70.6	R-Q	1272.8		
1966	R	29.8	122.2	164.7	301.3	150.6	198.4	131.6	152.3	156.6	88.0	132.4	2.9	1630.9			
	RD	7	15	15	25	19	21	17	15	19	17	14	2	186			
	Q	21.7	20.2	22.7	46.4	84.3	72.1	59.0	50.8	137.1	73.1	41.3	27.6			656.3	
	E	145.5	99.0	135.7	91.8	116.5	132.6	124.8	123.8	121.2	129.4	135.7	166.2				1522.2
	R ² Q	8.0	101.9	142.1	255.0	66.3	126.3	72.6	101.5	19.5	14.9	91.1	-24.6	R-Q	974.6		
1967	R	5.9	32.0	183.9	248.1	355.4	217.7	157.3	245.4	210.6	178.6	214.4	116.8	2166.3			
	RD	3	6	17	24	29	26	23	24	23	28	27	13	243			
	Q	17.9	12.0	13.3	18.5	47.7	64.2	87.1	82.3	86.9	84.3	80.5	126.7			721.4	
	E	190.4	170.2	163.8	121.5	100.1	107.1	104.6	111.3	115.2	108.5	103.2	142.8				1538.8
	R ² Q	-12.0	20.0	170.6	229.7	307.7	153.5	70.3	163.1	123.7	94.3	134.0	-9.9	R-Q	1444.8		

1968	R	18.4	239.3	180.5	340.6	218.0	260.1	222.9	303.0	218.8	164.0	172.1	108.1	2445.9	
	RD	5	23	18	27	27	27	26	25	21	23	20	12	254	
	Q	60.8	36.2	35.5	77.3	162.3	129.9	133.8	160.5	139.5	85.6	66.5	86.5	1174.4	
	E	175.5	117.0	133.2	97.7	102.5	93.6	93.8	100.1	124.8	116.2	110.7	125.9		1391.1
	R ² Q	-42.4	203.1	145.0	263.3	55.7	130.1	89.1	142.5	79.3	78.4	105.6	21.6	R-Q	1271.5
1969	R	107.7	228.0	172.7	101.0	253.6	152.5	155.8	194.5	112.8	129.4	68.0	9.1	1685.2	
	RD	16	13	14	16	25	18	22	23	13	20	16	7	203	
	Q	48.5	36.8	32.7	29.6	41.7	47.5	50.5	67.5	60.4	42.1	29.2	20.4	506.8	
	E	142.7	118.6	155.2	143.0	100.3	110.9	95.3	113.4	128.7	131.3	121.2	144.0		1504.7
	R ² Q	59.2	191.3	140.1	71.4	211.9	105.0	105.4	127.0	52.4	87.3	38.8	-11.3	R-Q	1178.4
1970	R	287.7	147.7	269.2	221.2	368.8	197.6	118.9	294.7	291.5	207.2	116.4	78.4	2599.2	
	RD	24	13	25	22	26	20	20	25	24	24	14	12	249	
	Q	25.8	21.2	37.5	67.3	172.3	111.5	92.3	84.5	144.2	135.7	86.7	48.0	1026.9	
	E	123.4	139.2	129.4	118.5	106.7	107.2	108.0	96.3	104.8	120.7	132.9	147.9		1432.9
	R ² Q	261.8	126.5	231.7	153.8	196.6	86.1	26.6	210.2	147.4	71.5	29.7	30.4	R-Q	1572.3
1971	R	91.1	8.8	47.6	275.5	355.6	367.8	283.7	370.0	175.4	145.2	104.3	138.0	2363.0	
	RD	10	2	6	24	29	26	22	28	24	20	17	15	223	
	Q	30.4	17.9	14.6	19.9	35.4	126.6	158.7	216.5	190.6	82.5	45.5	34.1	972.8	
	E	150.5	158.6	183.8	98.8	91.6	94.0	103.7	110.6	108.6	121.6	135.3	136.5		1493.6
	R ² Q	60.7	-9.1	33.0	255.6	320.2	241.2	125.0	153.5	-15.3	62.7	58.8	103.8	R-Q	1390.2
1972	R	51.5	125.7	85.3	166.0	332.1	275.2	230.9	213.9	139.7	215.2	229.4	115.8	2180.6	
	RD	9	15	12	23	28	24	28	22	18	28	22	12	241	
	Q	24.8	18.9	17.5	16.7	33.9	61.2	93.7	102.9	91.0	56.8	103.9	87.0	708.2	
	E	167.1	121.1	155.5	122.8	101.6	99.1	105.0	116.0	120.3	106.5	104.5	143.1		1462.7
	R ² Q	26.8	106.8	67.8	149.3	298.2	214.0	137.3	110.9	48.6	158.4	125.5	28.8	R-Q	1472.3
1973	R	146.3	237.6	14.4	155.2	341.2	240.7	150.7	229.6	202.4	168.0	89.0	37.7	2012.7	
	RD	15	18	6	18	25	26	14	27	22	23	14	7	215	
	Q	48.3	37.9	38.2	26.6	36.3	88.0	83.1	67.2	98.9	79.3	55.1	37.6	696.4	
	E	151.3	137.2	186.0	131.8	101.2	101.3	108.3	108.0	108.6	116.1	118.9	148.9		1517.6
	R ² Q	97.9	199.7	-23.8	128.6	304.9	152.7	67.7	162.4	103.5	88.7	33.8	.1	R-Q	1316.3
MEAN	R	87.0	113.9	160.3	255.7	279.2	199.2	175.0	219.7	199.0	171.0	164.3	104.7	2129.0	
	RD	11	12	15	22	25	23	21	24	22	23	19	13	236	
	Q	36.9	22.8	22.4	34.5	81.1	77.1	79.3	80.0	93.3	78.8	70.1	63.7	740.0	
	E	153.8	138.8	153.7	114.4	107.2	106.6	104.2	108.5	114.8	117.3	116.1	137.8		1473.2
	R ² Q	50.1	91.1	137.9	221.2	198.1	122.1	95.7	139.7	105.7	92.2	94.2	41.0	R-Q	1389.0
S.D.	R	77.5	77.7	68.1	83.9	76.9	67.4	55.5	70.1	70.9	37.9	106.3	73.2	339.2	
	RD	6	5	5	3	3	3	4	3	4	3	4	5	23	
	Q	29.4	12.4	10.1	21.9	62.9	44.1	48.6	51.7	46.5	36.8	46.2	56.8	288.2	
	E	17.5	19.3	17.5	14.2	8.7	10.8	10.0	8.3	8.5	11.3	16.1	16.5		40.3
	R ² Q	76.9	73.8	69.3	72.6	89.5	55.3	29.6	54.0	67.7	38.9	71.9	52.3	SD R-Q	195.7

R is monthly total of rainfall
RD is the number of days on which there was rain
Q is the monthly total of streamflow
E is the monthly total of the Penman open water evaporation index
R²Q is the total monthly change in soil moisture and groundwater storage, and actual evapotranspiration loss
S.D. is the standard deviation of the above variables for the period 1.58 to 12.73

Table 8.6

Kenyan montane rainforest catchment daily frequency distributions 1.64 to 12.73

INTERVAL	RAIN				FLOW				EVAP			
	Interval 6.6 mm				Interval 3.6 mm				Interval 0.6 mm			
	mm	%	Daily	%	mm	%	Daily	%	mm	%	Daily	%
1	5607.6	16.46	2015	53.36	7203.9	60.84	4930	84.36	.4	.00	1	.02
2	8150.8	23.93	853	22.59	4100.5	34.63	852	14.58	13.8	.06	14	.24
3	7421.1	21.79	460	12.18	459.1	3.88	55	.94	151.9	.64	97	1.66
4	5093.0	14.95	223	5.91	77.2	.65	7	.12	801.9	3.40	375	6.42
5	3860.5	11.33	131	3.47	.0	.00	0	.00	2146.2	9.11	790	13.52
6	1972.9	5.79	55	1.46	.0	.00	0	.00	3525.7	14.96	1066	18.24
7	972.6	2.86	23	.61	.0	.00	0	.00	4001.0	16.97	1029	17.61
8	302.6	.89	6	.16	.0	.00	0	.00	4045.6	17.16	902	15.43
9	330.8	.97	6	.16	.0	.00	0	.00	3499.1	14.84	689	11.79
10	126.9	.37	2	.05	.0	.00	0	.00	2533.5	10.75	444	7.60
11	.0	.00	0	.00	.0	.00	0	.00	1804.5	7.66	287	4.91
12	.0	.00	0	.00	.0	.00	0	.00	787.3	3.34	115	1.97
13	.0	.00	0	.00	.0	.00	0	.00	220.9	.94	30	.51
14	.0	.00	0	.00	.0	.00	0	.00	40.1	.17	5	.09
15	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
16	105.2	.31	1	.03	.0	.00	0	.00	.0	.00	0	.00
17	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
18	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
19	120.1	.35	1	.03	.0	.00	0	.00	.0	.00	0	.00
20	.0	.00	0	.00	.0	.00	0	.00	.0	.00	0	.00
TOTALS	34064.0		3776		11840.7		5844		23571.9		5844	

CHAPTER 9.

Comparison of the algorithms and catchment models.

9.1 Algorithms and data.

It would appear to be relatively easy to take the published code for an algorithm and mount it on a computer for compilation since one would expect it to be free from error. In the author's experience published code should always be treated with care as it may contain significant errors of coding, rely on non-standard features of FORTRAN, not be coded to do what it is said to be doing, or have features which the new compiler does not like such as having a conflicting mixture of real and integer numbers in expressions usually brought about by ignoring the 'name rule'.

The compiler used was the MICROSOFT FORTRAN Version 5.1, which proved to have problems in compilation that stemmed from the mixture of real and integer. There were the usual irrelevant error messages depending on the origin of the nest of errors. The solution to a code problem was hardly ever found by use of the elaborate 'CODE VIEW' (included as a tool with the compiler) as there were usually several 'depths' to a particular error. 'Code view' was, however, very useful in developing a new subroutine program. The ever reliable 'PRINT' command had to be used extensively to identify areas of code or subroutines where the problem originated: a problem area has been identified in the code when a second following 'PRINT' IS NOT REACHED when the program aborts.

The general approach to the code for a particular algorithm was as follows on the next page.

1. To study the program for anomalies and anachronisms and to update it with the latest 'top down' FORTRAN structures available if possible without adversely affecting the program concepts and logic.
2. To check that parameter and array names did not conflict with those in the model, data access, and model output analysis programs, and adjust the algorithm names .
3. To compile and build the algorithm into the code library as a subroutine to ensure complete compatibility of common blocks and subroutines.
4. To run the algorithm, and identify and correct problems such as code errors, array overflows, domain errors and computer numerical 'overflow' or 'underflow'.
5. To crudely flowchart the algorithm and decide if it is operating correctly in accordance with theory and the 'write up', and to decide whether modifications or additions are necessary.
6. To run the algorithm on the test data sets and modify the concepts, code, and algorithm control parameter values using the results of test runs.

Two libraries were developed, ROS.LIB and EVOLS.LIB, since the evolutionary algorithm and its subroutines in the second library combined as a .EXE file are as large as the rest in the same form:-

CONTROS.EXE	485,518 bytes	ROS.LIB	123,507 bytes
EVOL.EXE	485,542 bytes	EVOLS.LIB	53,343 bytes

The complexity of the two libraries are shown by the sections of the computer mapping with common blocks and subroutines:-

ROS.LIB - Library of common blocks and subroutines accessed by CONTROS.FOR the main control program.

A1.....AUG	AUG.....AUG
B1.....AUG	B10.....AUG
B2.....AUG	B3.....AUG
B4.....AUG	B5.....AUG
B6.....AUG	B7.....AUG
B9.....AUG	CON.....conja
CONJ.....conja	DS.....ds
EXTRA.....AUG	M7.....AUG
MAR.....MARQD	MARQ.....MARQD
MGENINVD.....marsubd	MODEL.....model3
MREDUCD.....marsubd	MROWD.....marsubd
MSUBD.....marsubd	MTRAND.....marsubd
OPTIR.....optir20	PRED.....PRED
READER.....READER	SCALE.....SCALD
SIMPLEX.....simp	STAT.....STAT
STEP.....simpsub	

EVOLS.LIB - Library of common blocks and subroutines accessed by **EVOL.FOR** the main control program.

A1.....model3	AUXFUN.....ROUT
B1.....reader	B10.....model3
B2.....reader	B3.....reader
B4.....ROUT	B5.....reader
B6.....stat	B7.....stat
B9.....model3	BLETAL.....ROUT
C1.....ROUT	CONSTR.....ROUT
DREHNG.....DREHNG	EXTRA.....model3
FUNCT1.....ROUT	GAUSSN.....gaussn

GNPOOL.....gnpool	KORR.....korr
M7.....model3	MINMAX.....ROUT
MODEL.....model3	MUTATE.....mutate
NEXGEN.....ROUT	OUTERR.....OUTERR
PARSEL.....ROUT	PIDATA.....gnpool
READER.....reader	SCALE.....SCALD
STAT2.....stat	TESCRI.....ROUT
TIMDEL.....mutate	

At first each algorithm was given its own library and control program, but this led to further errors 'back tracking' with corrections when these had to be made to each of the six libraries. One of the most difficult errors to deal with is the 'wandering error' caused by array overflow: problem areas are 'patched' to remove an apparent error and then a new error occurs elsewhere. An example of this was the array ERR(30) in MODEL which was dimensioned for the number of parameters instead of maximum days in the month. The missing element value for ERR(31) overlaid a section of the computer program mapping producing the apparent error.

The tests were then carried out on the three sets of data using the four, five and nine parameters shown by the values in the columns of Tables 9.2 - 9.20 at the end of this chapter: the optimized results of the four parameter set were carried over as fixed values for the set of five remaining parameters, and then both sets of results were allowed to 'float' to give the final optimized nine parameter set. The first two optimizations used the parameter initial starting points and limits proposed in the HYRRROM Package shown in Table 9.1 to

give a standard starting point for each set of data.

Table 9.1

Initial values of parameters as set in HYRROM package

Parameter Symbol	Computer Mnemonic	Initial value	Lower limit	Upper limit
P ₁	SS	3.1	0.0001	10.0
P ₇	RC	0.7	0.10	0.98
P ₁₂	RDEL	0.2	0.1	28.0
P ₁₀	RX	2.6	1.0	5.0
P ₁₁	RK	0.1	0.01	1.0
P ₅	FC	0.7	0.4	1.0
P ₂₁	GSU	90.4	70.0 45.2*	150.0 435.6*
P ₂₂	GSP	1.7	1.0	5.0
P ₂₃	GDEL	0.5	0.1	25.0

*Values used for Kenya forest trials.

At the top of each table of results the test derivations for algorithm control parameters are shown, and in the shaded area the observed data for the catchment is given as a check that the model has processed the observed data correctly for each run. F_0 is the 'without model' variance of the observed data, F is the minimum value of the objective function reached by the algorithm. The statistical terms have their normal

definitions as discussed earlier in this text.

9.2 Multimembered Evolution Strategy KORR.

This algorithm is based on the code given in Schwefel (1981) with the principal difference that the model parameters are scaled and bounded, and the random number generator program is seeded by hundredths of a second from the computer clock. The path is through GETTIM, with a time delay subroutine, TIMEDEL, to prevent the same seed being used in 20 succeeding calculations which take place in very much less time than 10^{-2} of a second. The code given by Schwefel at one point is simply 20 calculations of the same random number from the same seed, which is apt to have a rather small standard deviation when the next search vector is chosen!

Modifications to the main program EVOL were the introduction of parameter scaling, and the failure system based on the parameter IFAILS was modified to continue the search until convergence or the maximum number of iterations had been reached. The determination of the step size by $n^{-1/2}$ suggested by Schwefel, but not programmed, was tried and rejected in favour of a simple multiplier.

The main subroutine KORR again had the IFAILS system changed to be compatible with the control program EVOL, and program control variables throughout the subroutines had to be scaled to enable alternate paths through each subroutine and those called by it to be followed. Otherwise, the same path was followed each time and random changes ignored.

These modifications were repeated in seven of the subroutines given by Schwefel, and further changes made to the subroutines MODEL, READER, and STAT to make them compatible.

The size of arrays in common blocks were changed so that either the nine or thirty parameter MODEL could be interfaced into either library, EVOLS.LIB. or ROS.LIB.

9.3 The Rosenbrock algorithm.

This was the only algorithm initially available to the author which was actually working program code, but it worked extremely slowly taking some two to three hundred runs of the model function to optimize four parameters over a long time period. This could be as long as four or five hours.

The first modification was to change the data set to two blocks per month: a control block and the block containing the data for the month written by implied DO loops instead of direct DO loops. This reduced the time taken by over 80%.

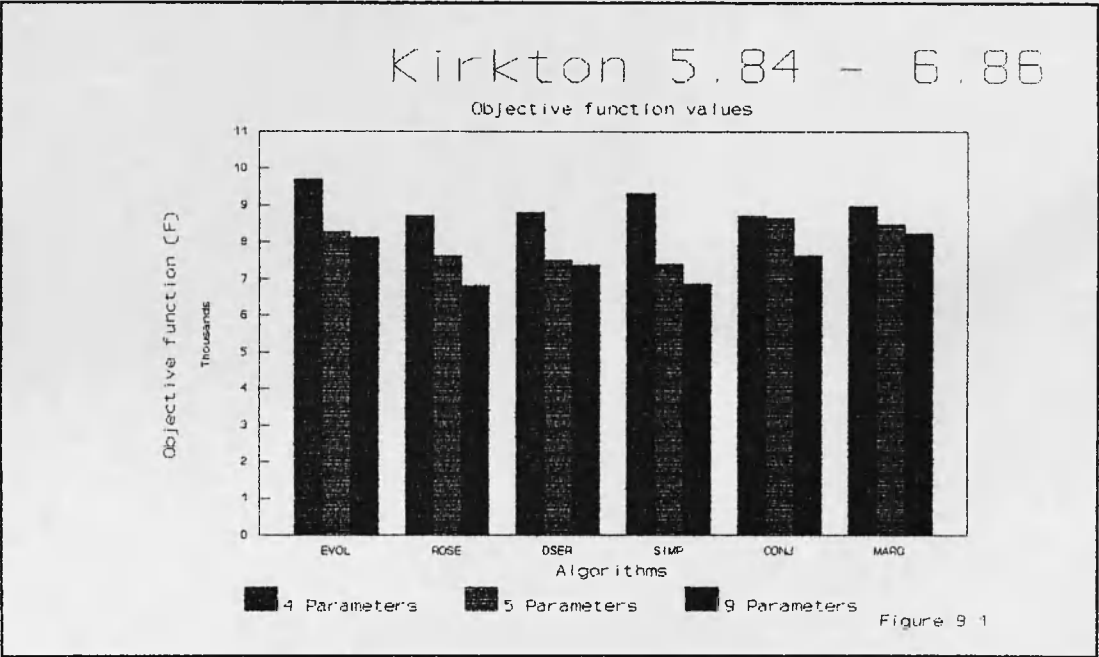
The next was to change the criterion for convergence from a preset change in each of the active parameter values, to a change in the explained variance or 'model fit'. This removed a large number of unnecessary trials which changed a parameter value, but due to its lack of sensitivity did not increase the explained variance.

Instead of using the original Rosenbrock criterion for termination of the search along a particular axis of ten successive failures, the algorithm had the 'success, success, failure, success' system built in to it together with quadratic interpolation. Coding errors had become part of the latter and these were identified and corrected.

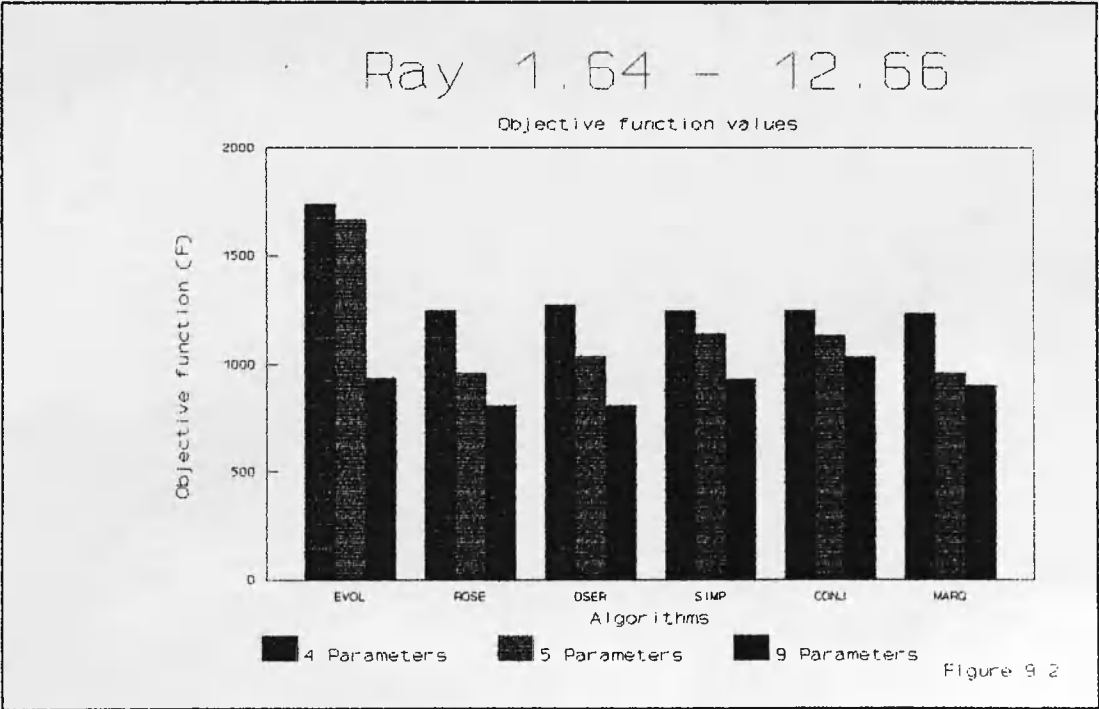
9.4 Direct Search algorithm.

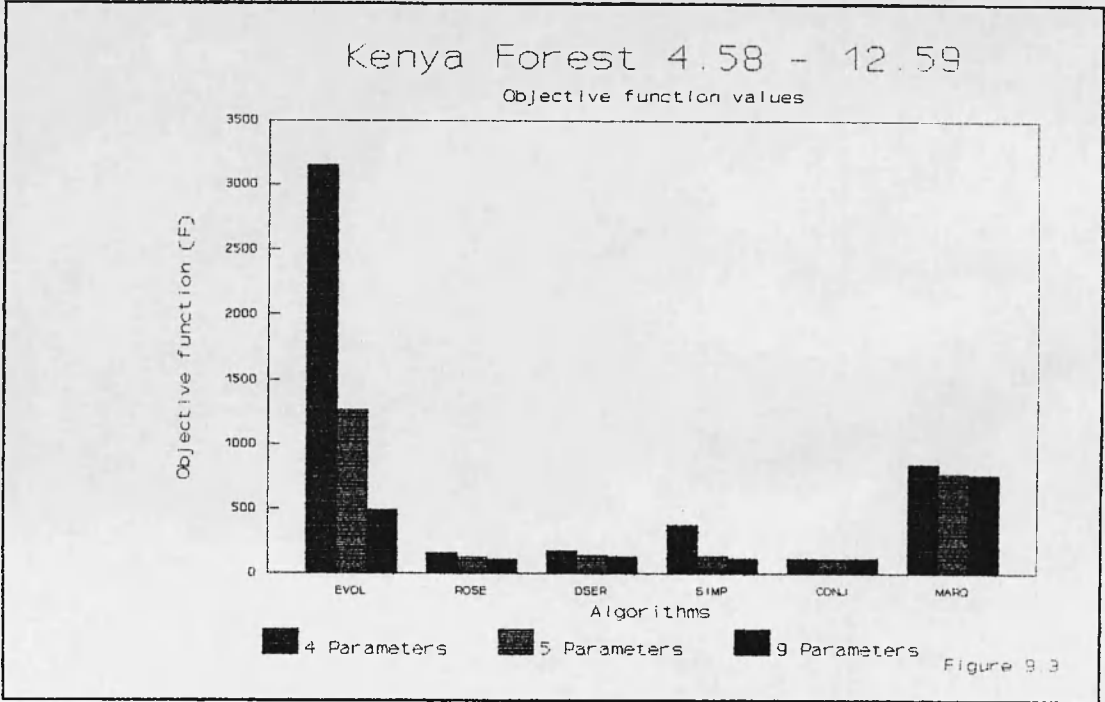
This algorithm was based on the initial part of the Rosenbrock algorithm by removing the Gram-Schmidt

orthogonalization program used for the Rosenbrock rotation of axes, but retaining the quadratic interpolation code. It



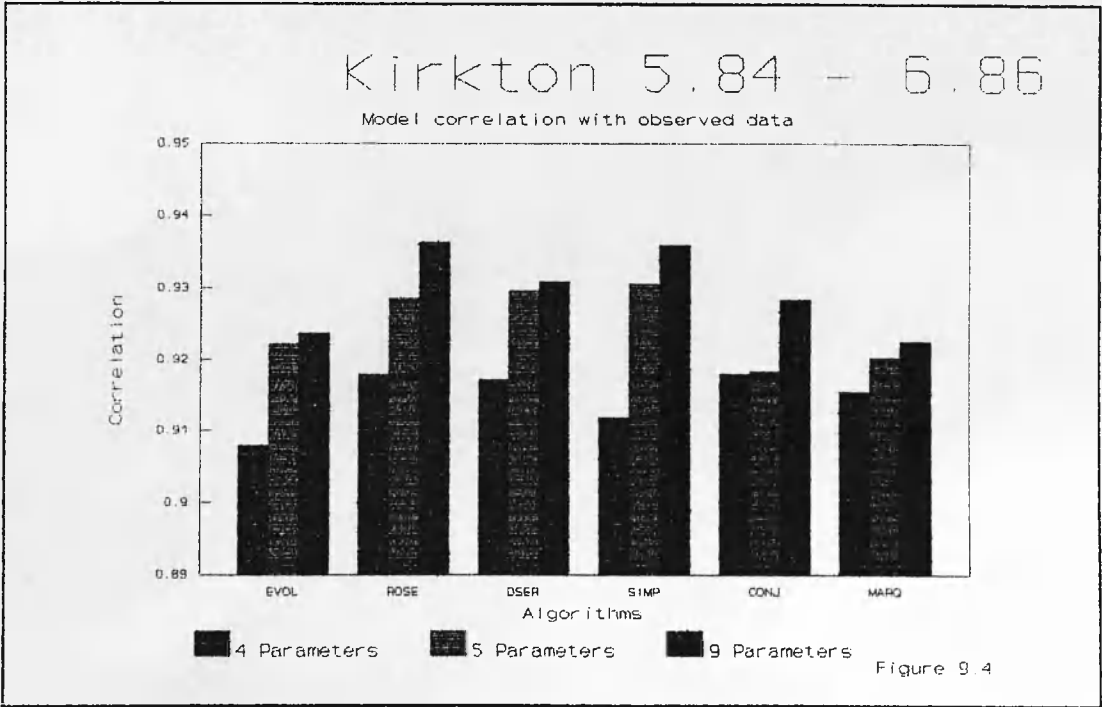
therefore presented a useful comparison of the difference made by using the Rosenbrock rotation of axes.



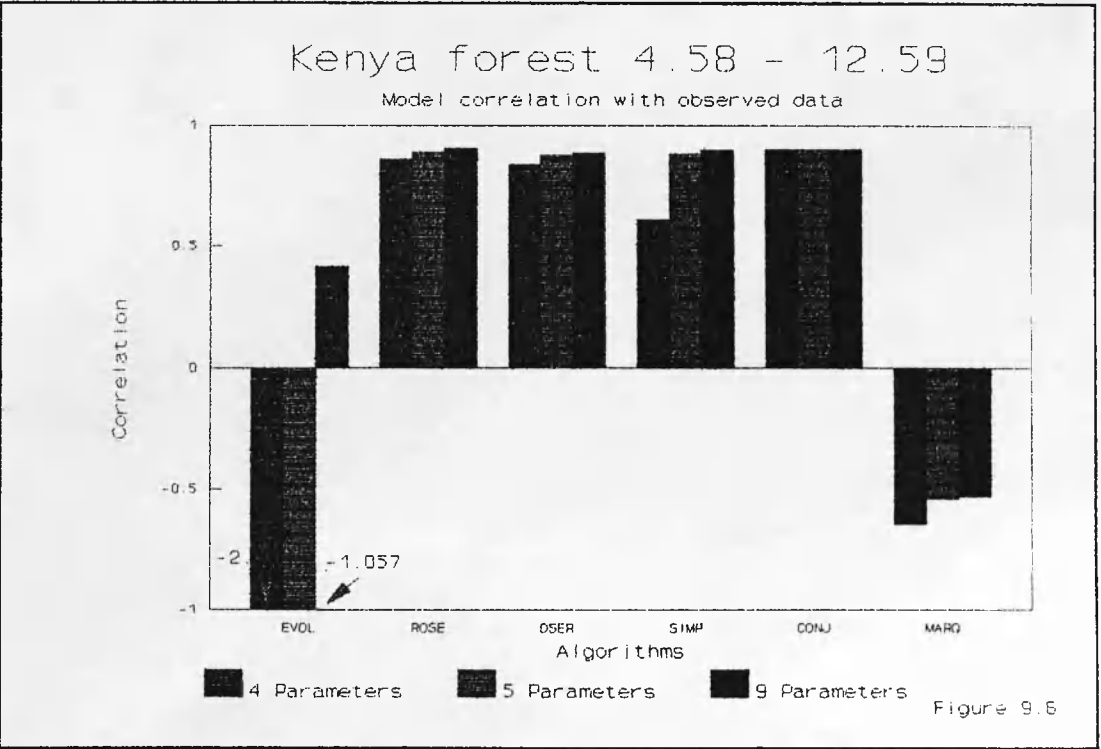
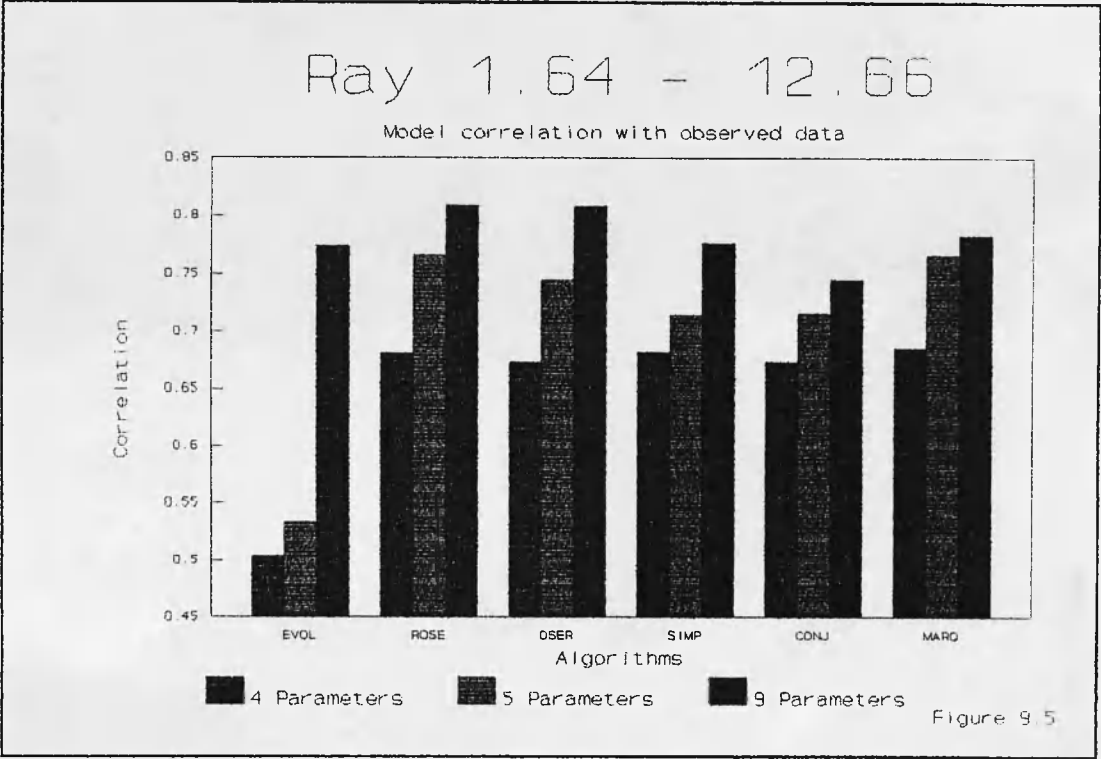


9.5 The Nelder and Mead Simplex algorithm.

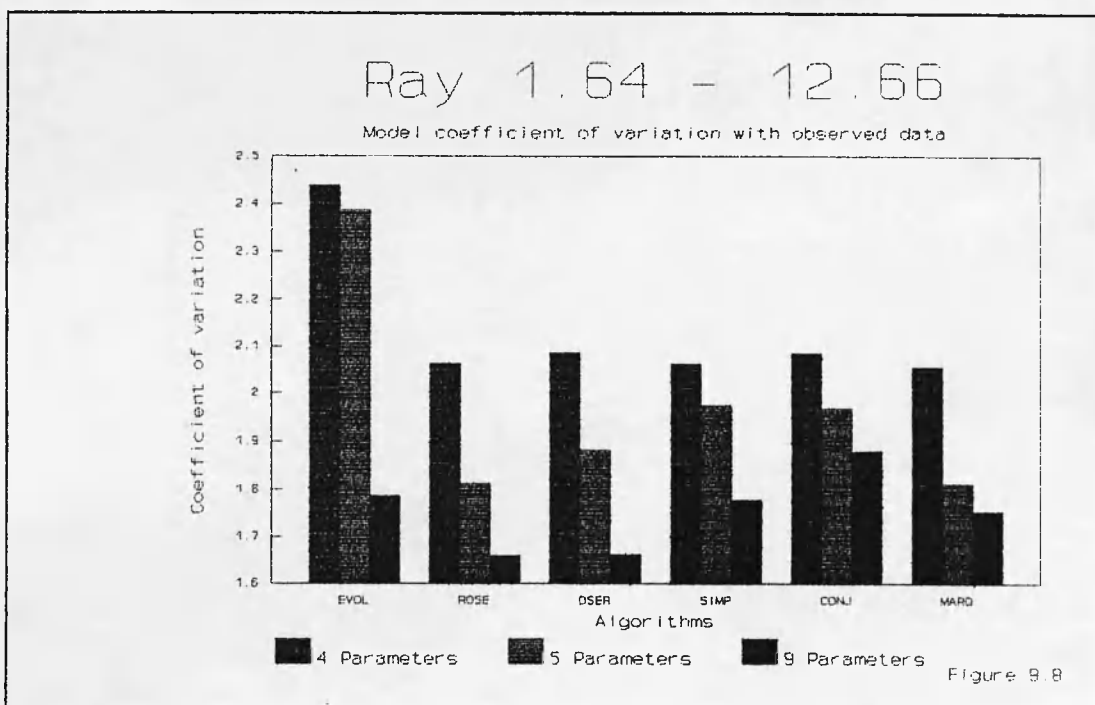
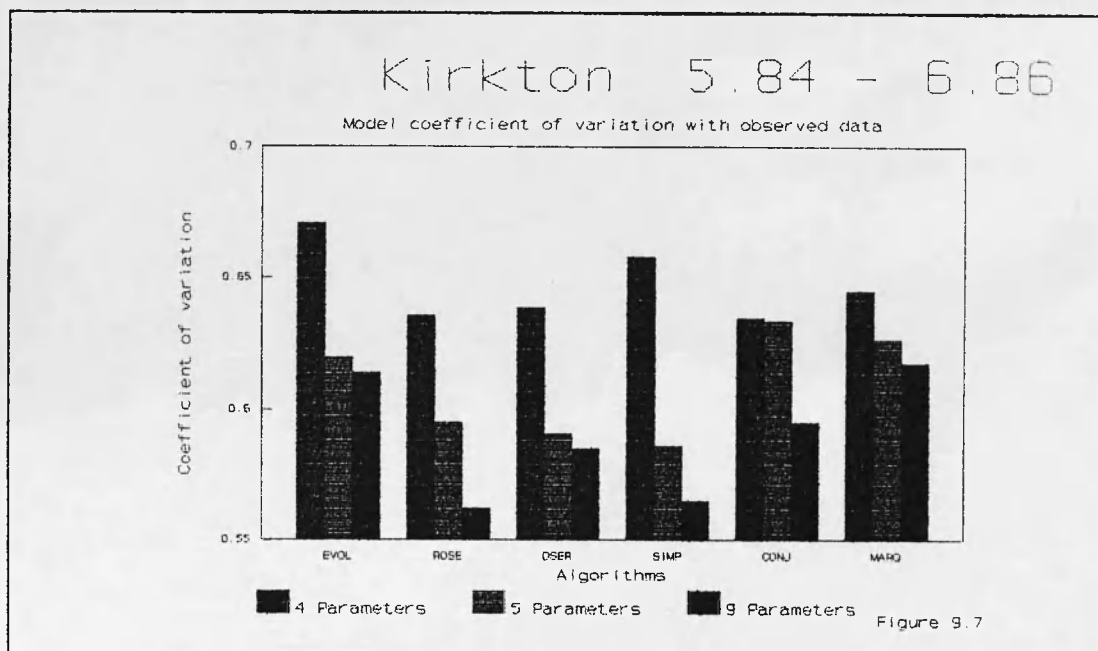
This algorithm is based on that presented in Clarke (1973) which definitely did not work when first mounted and tried on a computer although Clarke had found that it worked well when tested on the Rosenbrock parabolic valley. Problems were detected in the code which initially prevented the program



terminating normally as it became locked in program paths modifying the vertices, and values of control parameters were either very small or negative.



A new subroutine was developed modifying the simplex initialisation of step lengths so that the initial simplex

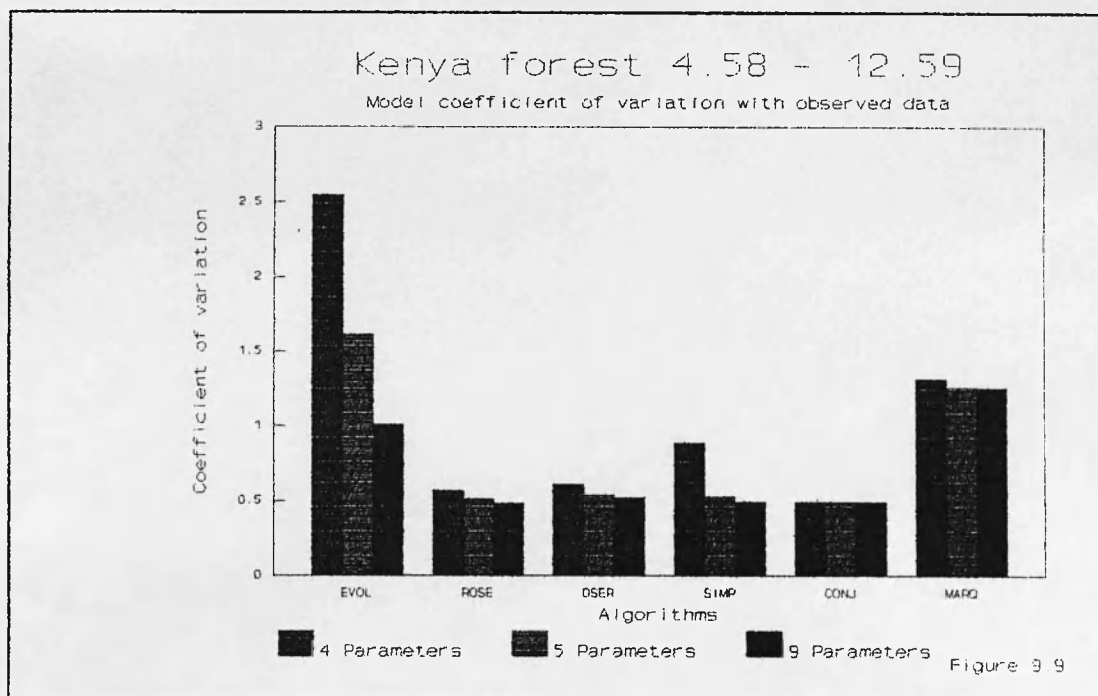


stayed in approximately the same position in n-dimensional space. The step length subroutine allowed tests to be made on the initialisation step multiplier for a complete new simplex, and a scheme of alternate positive and negative parameter steps

accelerated achievement of convergence. This depended on the closeness of the optimal point but usually gave a significant reduction in the number of model function calculations.

The modifications to the algorithm code greatly simplified the program reducing it by a third from the Clarke (1973) version, and removing computer generated domain and range errors preventing successful termination of the program.

Tests on the original three reflection, contraction and expansion coefficients proposed by Nelder and Mead produced no improvement on them in multiplier steps of 0.25, and they seem to be generally applicable with values of 1.0, 2.0 and 0.25. A rough test for global convergence was used by expanding the final simplex twice by a factor of 2.



9.6 The conjugate direction algorithm.

This is a development of the Harwell Subroutine Library VA04 integrated with the model and its subroutines. It needed

to have its program code adjusted and array sizes dimensioned correctly for the modelling problem. It also had domain problems and MATH errors, but did not require a great deal of effort to bring it up as a working program.

The program parameters were tested and modified but the really major difference between the recommended value for ESCALE of 100.0 and the best value found in the trials of 0.5! At the recommended value it prevented the operation of the conjugate direction part of the program by causing domain error problems. This parameter is used in the program as a model parameter weighting and forms a crude parameter bound. The model parameter scaling was not used, and the parameters were allowed to float without limits because of ESCALE. The use of a single column array, W, as workspace made it difficult to follow the working of the program. As well as being a control parameter for printing, IPRINT also defines paths through the program.

9.7 The Marquardt algorithm.

The basic code for this algorithm again came from the Harwell Subroutine Library and it was found necessary to use the double precision version VA05AD together with the algorithm subroutine MB11AD. Scaling and limits were introduced for the model parameters.

Problems again arose with the size of arrays; one having to be redefined as an n by n array from a single element array. The work space single column array, W, was split into two arrays in the algorithm subroutines; WB being used to store columns involved in matrix inversions, and W the original matrices, updated matrices and changes in parameters. This

division was necessary as sections of the array W were being incorrectly overwritten in the five algorithm subroutines. The use of a long, single column array for a workspace makes it difficult to follow the working of the program and it would be better if the array was split into separate arrays.

IPRINT, defined as a print control parameter, was again used as a path control through the program, and this time also changed its value within the program. The program parameter, H, used in estimating partial derivatives from difference approximations was found to have changed its name from H to DSTEP in the program which caused some problems with the input argument list.

The optimal value for DSTEP after 50 optimization runs with different values was found to be 1.44. However, it was also found that after the first iteration if the next step was preceded by the scaling of DSTEP by 10^{-1} that this enabled the Marquardt correction factor to become effective. The scaled parameters must be of the same magnitude to estimate the function partial derivatives from a difference approximation:-

$$\frac{\partial f_i}{\partial x_j} = \frac{[f_i(x_1, x_2, \dots, x_j + H, \dots, x_n) - f_i(x_1, x_2, \dots, x_j, \dots, x_n)]}{H}$$

where $H = DSTEP$

(9.1)

The x_j are the n scaled active parameters for the i-th function calculation. The use of DSTEP in two different roles: initially as a step change in the scaled parameter, and then in the above equation to determine the partial derivatives led to this scaling.

9.8 Results from the tests.

There are twenty tables (9.2 - 9.26) at the end of this Chapter which record the results of nearly four thousand model function calculations; these do not include the optimization runs used to find the 'best' values of algorithm parameters.

The data sets used were the Kirkton catchment, Ray catchment, and the Kenyan montane rain forest in increasing order of difficulty of optimization. The first set presented no great problems being well matched to the initial starting point parameters, but the second did not respond well because of problems with the loss functions caused by the upper limit set on the evapotranspiration parameter p_5 (F) of 1.0. This also affected the surface evaporation parameter p_3 (FS) from the equation:-

$$P_3 = 1 * P_5 \quad (9.2)$$

This equation is an approximation reducing the number of parameters from 10 to 9 in the smaller model. Without the limit on p_5 restricting the feasible space search the only unbounded algorithm, conjugate directions, produced a very small volume error, Table 9.15, compared to the others. Both the other sets of data are seasonally responsive but the Ray river usually dries out during the summer. The zero flow is very difficult to model unless the simulated flow is set to nothing when this occurs!

The third data set, Kenyan montane rain forest, has a problem caused by the very long baseflow recession as investigated in Chapter 8, section 8.8, and the upper limit on

the 'choke' parameter, p_{21} (GSU), was therefore extended from 150 to 435.

The Evolution Strategy, Tables 9.2 - 9.4, took the full 20 iterations allowed except for the last trial in 9.4, and producing good correlations for the first three trials on the Kirkton data. The Ray data produced a less successful fit with a bad simulated flow volume error, which was repeated on the Kenyan data with negative correlations on the four and five parameter trials.

The Rosenbrock algorithm, Tables 9.5 - 9.7, produced the best correlations with excellent minimal volume errors for the Kirkton data. It too found trouble dealing with the problems of the Ray data, but moved from a bad volume error on the independent trial with four parameters to quite a good one on the last dependent trial with the final correlation going above 90%.

The Direct Search algorithm, Tables 9.8 - 9.10, performed nearly as well as the Rosenbrock; not entirely unexpected as it is virtually the same algorithm without the rotation of axes. On the Kenya data it passed from a bad over prediction of volume to a bad under prediction in the final trial with a final correlation 12% less than the Rosenbrock. It did not pass beyond one iteration in its search for the optimum.

The Nelder and Mead Simplex, Tables 9.11 - 9.13, is the bounded one of the two unscaled algorithms. It took the full number of iterations, twenty, allowed on all data sets, and achieved good results on the Kirkton data comparable to the Rosenbrock results. It performed less well on the other two data sets but made a good attempt at reducing the volume error with nearly 90% correlation on the Kenya data.

The Powell conjugate directions algorithm, Tables 9.14 - 9.17, is the other unscaled program to the simplex, but has in addition no bounds set on the model parameters. The latter condition has allowed the algorithm to reach good volume fits, but has allowed the time delay to become a negative parameter for the Kenya data. It performed normally on the Kirkton data but did not optimize five parameters for the Ray and three for the Kenya data. It did, however, point to the need for a higher upper bound to the evapotranspiration factor, FC. Tables 9.16 and 9.17 contrast the value of DFS found by trials, 0.05, and the recommended value of 100 for the Kenya data. In the former case most of the progress was made using 153 function calculations for the four parameter trial and no progress with the final nine parameter trial. Only one iteration occurred for each trial, but better results were obtained than with DFS = 100 where the strategy worked for all trials but the same three parameters failed to optimize.

The Marquardt algorithm was very efficient in its use of function calculations only requiring ten or eleven for the final nine parameter trials. Again it performed well on the Kirkton data and not so well on the Ray data. It did not achieve a positive correlation on the Kenya data and failed to optimize three parameters. This result would seem to show that an algorithm based on finding slopes in the objective function space is very efficient in its use of model calculations, but that the slopes are not so well defined with a large data set and therefore the accuracy of its optimization is low.

9.9 Direct ranking of algorithms.

The algorithms are arranged in Figures 9.1 - 9.9 by test

order with histograms showing the four, five, and nine parameter trials, and each figure showing one of the three data sets' results for reduction of objective function, the increase in correlation of simulated data with observed data, and the decrease in the coefficient of variation.

The only complete failure to produce a positive correlation is with the Marquardt in Figure 9.6. The Evolution strategy produces the worst four parameter correlation but does manage to rectify its errors and produce a positive correlation by the nine parameter trial.

The algorithms are ranked in order in Table 9.2 by their relative performance in reducing the objective function, increasing the correlation and reducing the coefficient of variation.

The weighting number of the algorithms is the sum of their ranks in Table 9.22:-

Evolution strategy	= 16
Rosenbrock	= 3
Direct search	= 9
Simplex	= 9
Conjugate directions	= 12
Marquardt	= 14

9.10 Hybridization of two algorithms.

It is clear from the simple weighted rankings that the Rosenbrock algorithm is the best one used over the three data sets. Second place is shared by the Direct Search and the Simplex algorithms. However, Direct Search is the equivalent of rerunning the Rosenbrock algorithm and the author has found this to fail when near to an optimum - no further progress can

be made. This is particularly so when running an optimization on monthly totals for the volume response, and one on the daily totals for the rapid response parameters. Two different models develop and further progress cannot be made by optimization - the most 'realistic' parameter has to be chosen from each set and combined into a new set. This sometimes produces an immediate improvement in the reduction of the objective function, and further progress can be made by optimization. However, it was found by the author during various projects that using the Rosenbrock followed by the Simplex algorithm, e.g. Eeles and Blackie (1993), could produce better model parameters giving a lower F value than either of the single algorithms. This indicates that there is merit in combining the two together as a hybrid with the simplex operating when the search has been brought near to the optimum. Tests were made of combinations of the Rosenbrock with conjugate direction and the Marquardt algorithms; neither were found to be better than the Rosenbrock/Simplex algorithm.

A hybrid version of the two algorithms has been developed, and the results are shown in Tables 9.23 to 9.25 for each of the three data sets. The same test procedure as outlined in section 9.1 was applied for the four, five, and nine groups of parameters.

In Table 9.23 the results of the hybrid optimization for the Kirkton catchment are shown with the catchment observed data in the grey area at the top: although this repeats from Table to Table for each catchment it has proved a valuable check on the data sets and the working of the models. All the objective function (F) values are lower than those found in Table 9.5, although the volume error is higher in the nine

parameter test. GSP (p_{22}) stopped at its higher bound, 5.0, and this bound was allowed to go to 8.0 for the other two tests.

The Ray catchment results are shown in Table 9.24 and this time a lower bound was reached for GSU which was then lowered to 30.0 for the nine parameter test. The first two F-values are lower than those in Table 9.6 but the last test gives a slightly higher value. The volume errors are all lower although 12% is still not good. The improvement is caused by the higher evapotranspiration parameter, p_5 (FC), which is allowed to rise above its upper bound of 1.0 and increases the 'loss' function output.

The montane rain forest in Kenya results are shown in Table 9.25 with lower F-values and a better final volume error. The number of model runs are very much higher than with the Rosenbrock algorithm, but this is the penalty for greater accuracy and the extraction of more information from the data set.

9.11 Comparison of the two models.

The nine parameter model is designed to produce a fit to a calibration set of data with the objective of either infilling or extending a data set. The larger thirty parameter model is designed to make an estimate of the effects of land use change and allow the forecasting of its effects on river flow or catchment yield.

A comparison of the results of using the parameters from the Rosenbrock set, and the Hybrid set for the nine parameter model together with the thirty five parameter model are shown for the Kirkton calibration period in Table 9.26. The Kirkton is the only set of data for which there is a land use change

and details of the areas and times involved when forest clear felling took place. No calibration of the model was derived from the land use change period so heavy dependence had to be placed on the physical relevance of the empirical Calder and Newson (1979) equations. The loss function in the nine parameter model is the simple factor multiplier (FC) for the evapotranspiration data, while in the large model only the grassland loss function has this form. The other loss functions for heather and forest consist of one for interception and another for transpiration losses for each vegetation.

The correlations are surprising in their consistency considering the relative size of the models and may well be a comment on the amount of useful information that can be obtained by optimization from the observed data. Due to the inherent errors and precision of the observed data sets a model can never produce a perfect simulation so there must be a limit to the useful information.

A validation of each model by running it over the full set of data from 5.84 to 12.88 was made and the results compared and contrasted in Table 9.27. If the volume errors are 'corrected' for the calibration errors then the large model has a simulation error of -0.504%, the 'Rosenbrock' optimized model -1.030% and the 'Hybrid' optimized model -0.986%.

Table 9.2

Algorithm:-

EVOL (Scaled)

Parents 2, descendants 6, Rotation of hyperellipsoid - angle 60^0 (5^0 for 9 parameter trial), IRECOM 5, selection from parents and descendants.

Catchment	Kirkton	Total rain	4,997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
F_0	55,209.4	Total E_0	1,264.9 mm
F	9,705.6	8,267.2	8,117.2
No. model runs	160	160	160
Iterations	20	20	20
Exp. var.	0.8242	0.8503	0.8530
Correlation	0.9079	0.9221	0.9236
Coefficient of variation	0.671	0.620	0.614
Pred. flow	3,924.7 mm	3,925.8 mm	3,925.9 mm
Vol. error	-4.907%	-4.879%	-4.877%
Pred. AET	1,152.7 mm	1,152.7 mm	1,248.2 mm
Ratio AET/ E_0	0.911	0.911	0.987
SS	3.51291		4.40114
RC	0.56856		0.67145
RDEL		0.38807	0.38879
RX		2.82303	2.75508
RK		0.05523	0.09821
FC	0.74189		0.71800
GSU		95.83518	101.35560
GSP	2.12804		2.44635
GDEL		0.52223	0.83234

The computer mnemonics are described in Table 9.1, page 163.

Table 9.3

Algorithm:-

EVOL (Scaled)

Parents 2, descendants 6, Rotation of hyperellipsoid - angle 60° (5° for 9 parameter trial), IRECOM 5, selection from parents and descendants.

Catchment	Ray	Total rain	1,991.1 mm
Period	1.64 - 12.66	Total flow	566.0 mm
F_0	2,325.3	Total E_0	1,845.3 mm
F	1,736.8	1,664.8	932.3
No. model runs	160	160	126
Iterations	20	20	20
Exp. var.	0.2531	0.2840	0.5990
Correlation	0.5031	0.5329	0.7740
Coefficient of variation	2.438	2.387	1.786
Pred. flow	1,140.6 mm	1,161.8 mm	916.6 mm
Vol. error	101.536%	105.294%	61.946%
Pred. AET	329.7 mm	329.7 mm	875.3 mm
Ratio AET/ E_0	0.179	0.179	0.474
SS	0.59453		2.25141
RC	0.11613		0.35864
RDEL		0.47824	0.60654
RX		2.79701	3.42824
RK		0.07936	0.01400
FC	0.40011		0.68805
GSU		87.44703	86.45114
GSP	1.04422		1.01730
GDEL		0.27838	0.46896

The computer mnemonics are described in Table 9.1, page 163.

Table 9.4

Algorithm:-

EVOL (Scaled)

Parents 2, descendants 6, Rotation of hyperellipsoid - angle 60^0 (5^0 for 9 parameter trial), IRECOM 5, selection from parents and descendants.

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	3,155.3	1263.8	493.9
No. model runs	153	150	26
Iterations	20	20	5
Exp. var.	-4.286	-1.1175	0.1725
Correlation	-	-	0.4153
Coefficient of variation	2.544	1.610	1.006
Pred. flow	1,923.3 mm	1,317.2 mm	968.5 mm
Vol. error	244.278%	135.782%	73.373%
Pred. AET	734.9 mm	734.9 mm	840.7 mm
Ratio AET/ E_0	0.296	0.296	0.339
SS	1.55173		1.79730
RC	0.01012		0.01636
RDEL		0.01869	0.01327
RX		1.40601	0.93808
RK		0.00451	0.00052
FC	0.35897		0.84858
GSU		321.45770	349.54350
GSP	1.00110		1.00986
GDEL		6.05365	0.00412

The computer mnemonics are described in Table 9.1, page 163.

Table 9.5

Algorithm:-

ROSENBROCK (Scaled)

Algorithm with Rosenbrock rotation of axes, scaled parameters, quadratic interpolation (2.0) for axis lowest point, and termination of search when model fit is less than a predetermined level of change (0.5).

Catchment	Kirkton	Total rain	4997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
F_0	55,209.4	Total E_0	1,264.9 mm
F	8,706.2	7,616.7	6,807.8
No. model runs	37	47	79
Iterations	2	2	2
Exp. var.	0.8423	0.8620	0.8767
Correlation	0.9178	0.9285	0.9363
Coefficient of variation	0.636	0.595	0.562
Pred. flow	4,118.8 mm	4,120.3 mm	4,146.3 mm
Vol. error	-0.204%	-0.167%	0.462%
Pred. AET	557.2 mm	557.2 mm	492.7 mm
Ratio AET/ E_0	0.440	0.440	0.390
SS	1.16652		0.99344
RC	0.54029		0.62279
RDEL		0.34154	0.33208
RX		4.01925	1.29366
RK		0.76907	0.33208
FC	0.70569		0.70581
GSU		92.64974	106.08790
GSP	4.80172		4.75020
GDEL		0.12680	0.10262

The computer mnemonics are described in Table 9.1, page 163.

Table 9.6

Algorithm:-

ROSENBROCK (Scaled)

Algorithm with Rosenbrock rotation of axes, scaled parameters, quadratic function (2) for axis lowest point, and termination of search when model fit has reached a predetermined level of change (0.5).

Catchment	Ray	Total rain	1,991.1 mm
Period	1.64 - 12.66	Total flow	566.0 mm
F_0	2,325.3	Total E_0	1,845.3 mm
F	1,245.6	959.8	804.2
No. model runs	38	47	81
Iterations	2	2	2
Exp. var.	0.4643	0.5872	0.6541
Correlation	0.6814	0.7663	0.8088
Coefficient of variation	2.064	1.812	1.659
Pred. flow	878.6 mm	887.1 mm	837.1 mm
Vol. error	55.2%	56.7%	47.9%
Pred. AET	1,581.0 mm	1,581 mm	1,651.3 mm
Ratio AET/ E_0	0.857	0.857	0.895
SS	9.06672		8.55874
RC	0.25240		0.37528
RDEL		0.63039	0.58052
RX		2.84071	2.35814
RK		0.09566	0.10633
FC	0.70124		0.95582
GSU		71.23342	70.12093
GSP	1.70016		4.98351
GDEL		0.12224	0.10001

The computer mnemonics are described in Table 9.1, page 163.

Table 9.7**Algorithm:-****ROSENBROCK (Scaled)**

Algorithm with Rosenbrock rotation of axes, scaled parameters, quadratic function (2) for axis lowest point, and termination of search when model fit has reached a predetermined level of change (0.5).

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	156.7	126.941	111.0
No. model runs	36	61	85
Iterations	2	2	2
Exp. var.	0.7374	0.7873	0.8141
Correlation	0.8587	0.8873	0.9023
Coefficient of variation	0.567	0.510	0.477
Pred. flow	680.0 mm	535.0 mm	549.9 mm
Vol. error	21.733%	-4.237%	-1.572%
Pred. AET	1,609.5 mm	1,609.5 mm	1,514.9 mm
Ratio AET/ E_0	0.649	0.649	0.610
SS	3.682.83		3.43720
RC	0.08613		0.26987
RDEL		0.02806	0.10756
RX		1.01605	1.02672
RK		0.12894	0.02348
FC	2.47988		2.49966
GSU		317.38590	435.25520
GSP	1.00784		1.00194
GDEL		0.05009	0.14787

The computer mnemonics are described in Table 9.1, page 163.

Table 9.8

Algorithm:-

DIRECT SEARCH (Scaled)

Algorithm with scaled parameters, quadratic interpolation (2) for axis lowest point, and termination of search when model fit has reached a predetermined level of change (0.5).

Catchment	Kirkton	Total rain	4997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
F_0	55,209.4	Total E_0	1,264.9 mm
F	8,786.9	7,513.1	7,381.5
No. model runs	22	33	41
Iterations	1	1	1
Exp. var.	0.8408	0.8639	0.8663
Correlation	0.9170	0.9295	0.9308
Coefficient of variation	0.639	0.591	0.585
Pred. flow	4,064.5 mm	4,065.8 mm	4,123.4 mm
Vol. error	-1.519%	-1.489%	-0.092%
Pred. AET	714.2 mm	714.2 mm	428.2 mm
Ratio AET/ E_0	0.565	0.565	0.338
SS	1.61351		0.81417
RC	0.58399		0.59859
RDEL		0.36193	0.35159
RX		4.99782	1.61331
RK		0.91494	0.99295
FC	0.70942		0.77588
GSU		92.65300	93.24755
GSP	4.99264		4.74194
GDEL		0.15803	0.17540

The computer mnemonics are described in Table 9.1, page 163.

Table 9.9**Algorithm:-****DIRECT SEARCH (Scaled)**

Algorithm with scaled parameters, quadratic interpolation (2) for axis lowest point, and termination of search when model fit has reached a predetermined level of change (0.5).

Catchment	Ray	Total rain	1,991.1 mm
Period	1.64 - 12.66	Total flow	566.0 mm
F_0	2,325.3	Total E_0	1,845.3 mm
F	1,271.7	1034.2	806.3
No. model runs	20	30	54
Iterations	1	1	1
Exp. var.	0.4531	0.5552	0.6532
Correlation	0.6731	0.7451	0.8082
Coefficient of variation	2.086	1.881	1.661
Pred. flow	876.5 mm	886.0 mm	836.7 mm
Vol. error	54.9%	56.5%	47.8%
Pred. AET	1,597.3 mm	1,597.3 mm	1,711.3 mm
Ratio AET/ E_0	0.866	0.866	0.927
SS	9.45202		9.75796
RC	0.21458		0.37896
RDEL		0.60825	0.60825
RX		2.98073	2.54755
RK		0.10352	0.07976
FC	0.70000		0.97533
GSU		71.50586	70.10239
GSP	1.70000		4.99264
GDEL		0.12718	0.10184

The computer mnemonics are described in Table 9.1, page 163.

Table 9.10

Algorithm:-

DIRECT SEARCH (Scaled)

Algorithm with scaled parameters, quadratic interpolation (2) for axis lowest point, and termination of search when model fit has reached a predetermined level of change (0.5).

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	1,264.9 mm
F	178.4	139.4	131.5
No. model runs	33	26	45
Iterations	1	1	1
Exp. var.	0.7011	0.7665	0.7798
Correlation	0.8373	0.8755	0.8830
Coefficient of variation	0.605	0.535	0.519
Pred. flow	699.8 mm	547.0 mm	481.1 mm
Vol. error	25.271%	-2.090%	-13.885%
Pred. AET	1,554.6 mm	1,554.6 mm	1,554.6 mm
Ratio AET/ E_0	0.626	0.626	0.626
SS	3.53949		3.53949
RC	0.07456		0.07711
RDEL		0.20000	0.20000
RX		2.60000	1.12438
RK		0.10000	0.07488
FC	2.47831		2.49967
GSU		391.75860	435.31550
GSP	1.77733		1.05898
GDEL		0.04329	0.01225

The computer mnemonics are described in Table 9.1, page 163.

Table 9.11

Algorithm:-

SIMPLEX (unscaled)

Algorithm with alternate step signs in subroutine, valley search when convergence has occurred, and standard parameters: reflection coefficient = 1.0, expansion coefficient = 2.0, and reduction coefficient = 0.25. For convergence standard error of vertices must be less than 1.1.

Catchment	Kirkton	Total rain	4997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
E_0	55,209.4	Total E_0	1,264.9 mm
F	9,322.8	7,404.6	6,865.2
No. of model runs	72	68	87
Iterations	20	20	20
Exp. var.	0.8311	0.8659	0.8757
Correlation	0.9117	0.9305	0.9358
Coefficient of variation	0.658	0.586	0.565
Pred. flow	3,938.5 mm	3938.6 mm	3,938.2 mm
Vol. error	-4.571%	-4.571%	-4.579%
Pred. AET	1,111.0 mm	1,111.0 mm	1,114.3 mm
Ratio AET/ E_0	0.878	0.878	0.881
SS	3.24154		3.26488
RC	0.55665		0.62586
RDEL		0.37663	0.37591
RX		1.73668	1.70057
RK		0.09217	0.09013
FC	0.73430		0.73311
GSU		79.10110	79.34718
GSP	2.90168		3.95950
GDEL		0.42457	0.43240

The computer mnemonics are described in Table 9.1, page 163.

Table 9.12

Algorithm:-

SIMPLEX (unscaled)

Algorithm with alternate step signs in subroutine, valley search when convergence has occurred, and standard parameters: reflection coefficient = 1.0, expansion coefficient = 2.0, and reduction coefficient = 0.25. For convergence the standard error of the vertices must be less than 1.1.

Catchment	Ray	Total rain	1,991.1 mm
Period	1.64 - 12.66	Total flow	566.0 mm
F_0	2,325.3	Total E_0	1,845.3 mm
F	1,243.5	1,139.2	925.0
No. of model runs	28	90	86
Iterations	20	20	20
Exp. var.	0.4625	0.5101	0.6022
Correlation	0.6821	0.7142	0.7760
Coefficient of variation	2.063	1.974	1.779
Pred. flow	885.7 mm	888.5 mm	901.4 mm
Vol. error	56.490%	56.996%	59.270%
Pred. AET	1,312.0 mm	1,312.0 mm	1,321.1 mm
Ratio AET/ E_0	0.711	0.711	0.716
SS	4.74453		4.85899
RC	0.28505		0.40289
RDEL		0.25009	0.44608
RX		2.23734	2.22896
RK		0.07139	0.06817
FC	0.73335		0.72160
GSU		80.30981	79.15796
GSP	1.59128		1.60964
GDEL		0.34314	0.33512

The computer mnemonics are described in Table 9.1, page 163.

Table 9.13

Algorithm:-

SIMPLEX (unscaled)

Algorithm with alternate step signs in subroutine, valley search when convergence has occurred, and standard parameters: reflection coefficient = 1.0, expansion coefficient = 2.0, and reduction coefficient = 0.25. For convergence the standard error of the vertices must be less than 1.1.

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	1.64 - 12.66	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	377.5	135.0	117.5
No. of model runs	54	49	90
Iterations	20	20	20
Exp. var.	0.3675	0.7738	0.8032
Correlation	0.6062	0.8796	0.8962
Coefficient of variation	0.880	0.526	0.491
Pred. flow	901.0 mm	632.0 mm	545.1 mm
Vol. error	61.288%	13.137%	-2.424%
Pred. AET	1,546.5 mm	1,546.5 mm	2,156.9 mm
Ratio AET/ E_0	0.623	0.623	0.869
SS	3.52069		5.20358
RC	0.13113		0.18593
RDEL		0.25616	0.27139
RX		1.00000	1.03104
RK		0.07147	0.05150
FC	1.69878		2.00494
GSU		435.6	435.59990
GSP	1.03408		1.02327
GDEL		0.38829	0.46724

The computer mnemonics are described in Table 9.1, page 163.

Table 9.14**Algorithm:-****Conjugate direction (unscaled)**

Modified version of Powell's algorithm with ESCALE = 0.5,
 ICONV = 1, and DFS = 0.05

Catchment	Kirkton	Total rain	4,997.2 mm
Period	05.84 - 06.86	Total flow	4,127.2 mm
F_0	55,209.4	Total E_0	1.265.0 mm
F	8,690.4	8,652.6	7625.9
No. model runs	38	33	58
Iterations	3	2	3
Exp. var.	0.8426	0.8433	0.8619
Correlation	0.9179	0.9183	0.9284
Coefficient of variation	0.635	0.634	0.595
Pred. flow	3,956.4 mm	3,954.2 mm	4,071.5 mm
Vol. error	-4.138%	-4.191%	-1.349%
Pred. AET	1,078.4 mm	1,078.4 mm	1,012.4 mm
Ratio AET/ E_0	0.852	0.852	0.800
SS	3.09770		3.09786
RC	0.54999		0.60047
RDEL		0.37168	0.34631
RX		3.60000	3.60000
RK		0.69910	0.16859
FC	0.69773		0.47442
GSU		90.3999	141.42510
GSP	7.10938		6.96422
GDEL		0.50000	0.07464

The computer mnemonics are described in Table 9.1, page 163.

Table 9.15

Algorithm:-

CONJUGATE DIRECTION (unscaled)

Modified version of Powell's algorithm with ESCALE = 0.5,
ICONV = 1, and DFS = 0.05.

Catchment	Ray	Total rain	1,991.1 mm
Period	1.64 - 12.66	Total flow	566.0 mm
E_0	2,325.3	Total E_0	1,845.3 mm
F	1,246.0	1,132.5	1,032.6
No. model runs	67	10	17
Iterations	1	1	1
Exp. var.	0.4534	0.5130	0.5559
Correlation	0.6734	0.7162	0.7456
Coefficient of variation	2.085	1.969	1.880
Pred. flow	577.5 mm	578.5 mm	590.9 mm
Vol. error	2.035%	2.212%	4.404%
Pred. AET	1.284.9 mm	1,269.4 mm	1,845.3 mm
Ratio AET/ E_0	0.696	0.688	0.688
SS	3.10647		3.10797
RC	0.27082		0.41851
RDEL		0.54941	0.54255
RX		3.60000	3.6
RK		0.1	0.1
FC	3.19637		3.19637
GSU		90.4	90.4
GSP	10.48538	1.7 (start)	1.7
GDEL		0.5	0.5

The computer mnemonics are described in Table 9.1, page 163.

Table 9.16

Algorithm:-

CONJUGATE DIRECTION (unscaled)

Modified version of Powell's algorithm with ESCALE = 0.5,
ICONV = 1, and DFS = 0.05.

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	115.5	114.4	114.4
No. model runs	153	9	7
Iterations	1	1	1
Exp. var.	0.8064	0.8083	0.8083
Correlation	0.8980	0.8990	0.8990
Coefficient of variation	0.487	0.484	0.484
Pred. flow	523.0 mm	523.1 mm	523.1 mm
Vol. error	-6.376%	-6.363%	-6.363%
Pred. AET	2,851.9 mm	2,851.9 mm	2,851.9 mm
Ratio AET/ E_0	1.149	1.149	
SS	7.41183		
RC	0.04825		
RDEL		-0.55887	
RX		2.11112	
RK		0.1	
FC	2.69970		
GSU		90.40000	
GSP	0.21151		
GDEL		0.50000	

The computer mnemonics are described in Table 9.1, page 163.

Table 9.17

Algorithm:-

CONJUGATE DIRECTION (unscaled)

Modified version of Powell's algorithm with ESCALE = 0.5,
ICONV = 1, and DFS = 100.0.

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	176.6	175.4	171.4
No. model runs	53	29	10
Iterations	2	1	1
Exp. var.	0.7041	0.7061	0.7129
Correlation	0.8391	0.8403	0.8443
Coefficient of variation	0.602	0.600	0.593
Pred. flow	400.7 mm	400.8 mm	429.4 mm
Vol. error	-28.264%	-28.262%	-23.141%
Pred. AET	2,676.6 mm	2,676.6 mm	2,533.6 mm
Ratio AET/ E_0	1.078	1.078	1.021
SS	6.75523		6.30506
RC	0.11841		0.12199
RDEL		-0.01500	-0.01500
RX		2.50148	2.50145
RK		0.10000	0.10000
FC	7.61609		7.61609
GSU		90.4	90.40000
GSP	8.04755		8.04755
GDEL		0.50000	0.50000

The computer mnemonics are described in Table 9.1, page 163.

Table 9.18**Algorithm:-****MARQUARDT (scaled)**

Modified double precision version of Powell's algorithm with
 MAXFUN = 200, IPRINT = 1, ACC = 0.05D0, DMAX = 0.8D0,
 DSTEP = 1.44D0.

Catchment	Kirkton	Total rain	4,997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
F_0	55,209.4	Total E_0	1,265.0 mm
F	8,957.7	8,467.0	8,237.7
No. model runs	7	8	11
Iterations	3	3	2
Exp. var.	0.8378	0.8466	0.8508
Correlation	0.9153	0.9201	0.9224
Coefficient of variation	0.645	0.627	0.618
Pred. flow	3,926.3 mm	3,926.9 mm	3,970.6 mm
Vol. error	-4.868%	-4.854%	-3.794%
Pred. AET	1,065.3 mm	1,065.3 mm	1,079.5 mm
Ratio AET/ E_0	0.842	0.842	0.853
SS	2.93942		3.14058
RC	0.50653		0.52518
RDEL		0.24632	0.26744
RX		3.73061	2.69151
RK		0.95287	0.96141
FC	0.78781		0.67664
GSU		93.56322	95.12463
GSP	4.68708		4.73081
GDEL		0.60074	0.65449

The computer mnemonics are described in Table 9.1, page 163.

Table 9.19

Algorithm:-

MARQUARDT (scaled)

Modified double precision version of Powell's algorithm with
 MAXFUN = 200, IPRINT = 1, ACC = 0.05D0, DMAX = 0.8D0,
 DSTEP = 1.44D0.

Catchment	Ray	Total rain	1,991.1 mm
Period	01.64 - 12.66	Total flow	566.0 mm
F_0	2,325.3	Total E_0	1,845.3 mm
F	1,233.7	959.6	898.8
No. model runs	6	7	11
Iterations	2	3	4
Exp. var.	0.4694	0.5873	0.6135
Correlation	0.6852	0.7664	0.7833
Coefficient of variation	2.055	1.812	1.754
Pred. flow	919.6 mm	917.2 mm	886.5 mm
Vol. error	62.480%	62.054%	56.634%
Pred. AET	1,426.0 mm	1,426.0 mm	1,488.0 mm
Ratio AET/ E_0	0.773	0.773	0.806
SS	6.29166		6.74975
RC	0.30376		0.33997
RDEL		0.51158	0.59009
RX		3.01726	3.19731
RK		0.16855	0.20257
FC	0.67962		0.77101
GSU		98.17321	101.66240
GSP	4.14392		4.28685
GDEL		0.74566	0.86942

The computer mnemonics are described in Table 9.1, page 163.

Table 9.20**Algorithm:-****MARQUARDT (scaled)**

Modified double precision version of Powell's algorithm with
 MAXFUN = 200, IPRINT = 1, ACC = 0.05D0, DMAX = 0.8D0,
 DSTEP = 1.44D0.

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	845.4	770.5	763.6
No. model runs	5	7	10
Iterations	3	3	3
Exp. var.	-0.4164	-0.2909	-0.2794
Correlation	-0.6453	-0.5393	-0.5286
Coefficient of variation	1.317	1.257	1.251
Pred. flow	572.2 mm	517.2 mm	497.8 mm
Vol. error	2.427%	-7.413%	-10.894%
Pred. AET	3,820.1 mm	3,820.1 mm	3,820.1 mm
Ratio AET/ E_0	1.539	1.539	1.539
SS	14.02415		14.02415
RC	0.41065		0.41065
RDEL		0.2	0.2
RX		4.11047	2.60000
RK		0.96457	0.96457
FC	2.33061		2.33061
GSU		417.60160	417.60160
GSP	2.96511		1.70000
GDEL		29.99979	0.50000

The computer mnemonics are described in Table 9.1, page 163.

Table 9.21

Ranking of algorithms for each data set.

Catchment	Algorithm in Test order	Statistic functions rank [*]	Algorithm by rank
Kirkton	Evolution Rosenbrock Direct srh. Simplex Conjugate Marquardt	5. 1. 3. 2. 4. 6.	Rosenbrock Simplex Direct srh. Conjugate Evolution Marquardt
Ray	Evolution Rosenbrock Direct srh. Simplex Conjugate Marquardt	5. 1. 2. 4. 6. 3.	Rosenbrock Direct srh. Marquardt Simplex Evolution Conjugate
Kenya forest	Evolution Rosenbrock Direct srh. Simplex Conjugate Marquardt	6. 1. 4. 3. 2. 5.	Rosenbrock Conjugate Simplex Direct srh. Marquardt Evolution

* Functions considered by the ranking are objective function, correlation, and coefficient of variation.

Table 9.22**Algorithm:-****HYBRID**

Hybridized optimization algorithm developed from a combination of the Rosenbrock and Simplex algorithms, with parameters bounded and scaled.

Catchment	Kirkton	Total rain	4997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
FO	55,209.4	Total EO	1,264.9 mm
F	8,683.4	7,522.5	6,746.6
No. model runs	83	97	190
Iterations	11	11	8
Exp. var.	0.8427	0.8637	0.8778
Correlation	0.9180	0.9294	0.9369
Coefficient of variation	0.635	0.591	0.560
Pred. flow	4,142.6 mm	4,144.0 mm	4,201.1 mm
Vol. error	0.371%	0.406%	1.790%
Pred. ET	499.3 mm	499.3 mm	367.1 mm
Ratio ET/EO	0.395	0.395	0.290
SS	1.01227		0.67802
RC	0.53649		0.65524
RDEL		0.31525	0.31338
RX		1.04668	1.28213
RK		0.87974	0.33104
FC	0.69659		0.67687
GSU		147.74980	111.27500
GSP	5.00000	7.15112	4.61235
GDEL		0.10003	0.19445

The computer mnemonics are described in Table 9.1, page 163.

Table 9.23

Algorithm:-

HYBRID

Hybridized optimization algorithm developed from a combination of the Rosenbrock and Simplex algorithms, with parameters bounded and scaled.

Catchment	Kirkton	Total rain	4997.2 mm
Period	5.84 - 6.86	Total flow	4,127.2 mm
F_0	55,209.4	Total E_0	1,264.9 mm
F	8,683.4	7,522.5	6,746.6
No. model runs	83	97	190
Iterations	11	11	8
Exp. var.	0.8427	0.8637	0.8778
Correlation	0.9180	0.9294	0.9369
Coefficient of variation	0.635	0.591	0.560
Pred. flow	4,142.6 mm	4,144.0 mm	4,201.1 mm
Vol. error	0.371%	0.406%	1.790%
Pred. AET	499.3 mm	499.3 mm	367.1 mm
Ratio AET/ E_0	0.395	0.395	0.290
SS	1.01227		0.67802
RC	0.53649		0.65524
RDEL		0.31525	0.31338
RX		1.04668	1.28213
RK		0.87974	0.33104
FC	0.69659		0.67687
GSU		147.74980	111.27500
GSP	5.00000	7.15112	4.61235
GDEL		0.10003	0.19445

The computer mnemonics are described in Table 9.1, page 163.

Table 9.24**Algorithm:-****HYBRID**

Hybridized optimization algorithm developed from a combination of the Rosenbrock and Simplex algorithms, with parameters bounded and scaled.

Catchment	Ray	Total rain	1,991.1 mm
Period	1.64 - 12.66	Total flow	566.0 mm
F_0	2,325.3	Total E_0	1,845.3 mm
F	1,166.0	883.8	808.2
No. model runs	69	57	164
Iterations	5	3	11
Exp. var.	0.4986	0.6199	0.6524
Correlation	0.7061	0.7873	0.8077
Coefficient of variation	1.997	1.739	1.663
Pred. flow	841.6 mm	844.8 mm	636.6 mm
Vol. error	48.701%	49.263%	12.479%
Pred. AET	1,845.3 mm	1,511.8 mm	2,048.4 mm
Ratio AET/ E_0	0.819	0.819	1.110
SS	6.19624		9.89896
RC	0.27952		0.44126
RDEL		0.60192	0.51069
RX		2.81659	2.37443
RK		0.08982	0.10060
FC	0.97963		2.32665
GSU		70.00000	47.37534
GSP	4.86778		6.59005
GDEL		0.10056	0.01816

The computer mnemonics are described in Table 9.1, page 163.

Table 9.25

Algorithm:-**HYBRID**

Hybridized optimization algorithm developed from a combination of the Rosenbrock and Simplex algorithms, with parameters bounded and scaled.

Catchment	Kenya forest	Total rain	3,321.0 mm
Period	4.58 - 12.59	Total flow	558.6 mm
F_0	596.8	Total E_0	2,481.8 mm
F	134.2	128.8	102.1
No. model runs	148	73	140
Iterations	21	11	21
Exp. var.	0.7751	0.7842	0.8290
Correlation	0.8804	0.8856	0.9105
Coefficient of variation	0.525	0.514	0.458
Pred. flow	442.5 mm	440.5 mm	558.8 mm
Vol. error	-20.795%	-21.148	0.030%
Pred. AET	2,297.7 mm	2,297.7 mm	1,797.1 mm
Ratio AET/ E_0	0.926	0.926	0.724
SS	5.59086		4.18410
RC	0.10578		0.10399
RDEL		0.19733	0.11563
RX		2.05141	1.68684
RK		0.12915	0.12584
FC	4.37296		4.82047
GSU		88.04596	84.59837
GSP	3.48967		3.01365
GDEL		5.59565	5.46959

The computer mnemonics are described in Table 9.1, page 163.

Table 9.26**Comparison of model calibrations**

The models used are:-

Column 1 - 9 Parameter model, Rosenbrock algorithm
 Column 2 - 9 Parameter model, Hybrid algorithm
 Column 3 - 30 Parameter model, Hybrid algorithm

Catchment	Kirkton	Total rain	3,737.1 mm
Period	5.84 - 12.85	Total flow	3,065.9 mm
F_0	40,792.7	Total E_0	958.0 mm
F	4,023.6	4,033.3	3,710.1
Column	1	2	3
Parameters	9	9	30
Exp. var.	0.9014	0.9011	0.9090
Correlation	0.9494	0.9493	0.9534
Coefficient of variation	0.511	0.512	0.491
Pred. flow	3,090.6 mm	3,113.1 mm	2,977.8 mm
Vol. error	0.807%	1.540%	-2.872%
Pred. AET	318.0 mm	273.5 mm	645.1 mm
Ratio AET/ E_0	0.332	0.285	0.673
SS	0.81833	0.67802	2.37693
RC	0.64682	0.65524	0.75592
RDEL	0.30329	0.31338	0.31316
RX	1.28321	1.28213	1.40563
RK	0.33199	0.33104	0.19103
FC	0.70573	0.67687	0.47485
GSU	105.00790	111.27500	62.41444
GSP	4.70117	4.61235	1.73874
GDEL	0.10292	0.19445	0.80850

The computer mnemonics are described in Table 9.1, page 163.

Table 9.27**Comparison of model validations**

The models used are:-

Column 1 - 9 Parameter model, Rosenbrock algorithm
 Column 2 - 9 Parameter model, Hybrid algorithm
 Column 3 - 30 Parameter model, Hybrid algorithm

Catchment	Kirkton	Total rain	10,918.4 mm
Period	5.84 - 12.88	Total flow	9,090.2 mm
F_0	117,633.4	Total E_0	2,523.8 mm
F	16,822.0	16,820.2	16,026.8
Column	1	2	3
Parameters	9	9	30
Exp. var.	0.8570	0.8570	0.8638
Correlation	0.9257	0.9258	0.9294
Coefficient of variation	0.589	0.589	0.491
Pred. flow	9,063.3 mm	9,140.6 mm	8,783.3 mm
Vol. error	-0.296%	0.554%	-3.376%
Pred. AET	948.5 mm	813.7 mm	1,839.8 mm
Ratio AET/ E_0	0.376	0.322	0.729
SS	0.81833	0.67802	2.37693
RC	0.64682	0.65524	0.75592
RDEL	0.30329	0.31338	0.31316
RX	1.28321	1.28213	1.40563
RK	0.33199	0.33104	0.19103
FC	0.70573	0.67687	0.47485
GSU	105.00790	111.27500	62.41444
GSP	4.70117	4.61235	1.73874
GDEL	0.10292	0.19445	0.80850

The computer mnemonics are described in Table 9.1, page 163.

Chapter 10.

Conclusions and recommendations for further work.

10.1 The current relevance of the research.

At the beginning of this research there were the first signs of the coming domination of scientific research by the 'commissioned research' project. It was still possible to take an investigative approach to a research topic without the possibility of an immediate cash return, but it was becoming a management preference that there should be a research sponsor with funds to devote to a specific objective. The commissioned project was supposed to have a significant 'research' content rather than be just a source of income. Since then the commissioned or consultant approach has become dominant with its 'short' cuts and 'deadlines' to be met, and its project research content has become very commercialised - all scientists have to earn their cost plus that of the overheads for the infrastructure.

The current state of the art of modelling was demonstrated at a British Hydrological Society meeting at Imperial College in June 1993 on 'Rainfall/Runoff Modelling'. Two members of the Institute of Hydrology, Tanya Jones and Frank Law, presented the paper 'A less complex package (HYRROM) for the rapid modelling of runoff time series', and a demonstration of the HYRROM package using a projection of the computer terminal screen.

Optimization was mentioned at various points during the discussion, but the only technique which appeared to be used successfully and consistently was the Rosenbrock algorithm.

The usual confusion existed between the 'reality' of a

model compared to the real environment, and it was thought that the values of model parameters could be established directly from the data or by field determinations beforehand. The only model labelled 'physically' based, SHETRAN (the former SHE with sediment transport added), had the claim made for it that it only needed field data to fix parameters to be able to forecast the effects of landuse change and climate change. The effective use of this model was then disposed of by its failure to perform well in two out of the four 'blind tests' where the model is set up with field determined parameters, and has to meet four sets of 'range' criteria in relation to observed streamflow.

Professor Keith Bevan of the University of Lancaster in the last paper of the meeting brought out the main problem of all current models that they have multiple optimal solutions to a single problem, equifinality in hydrological modelling, and that there was no unique parameter set. How could the discipline of rainfall/runoff modelling progress when this condition prevails?

The fact that a model is an abstraction of a 'simple' system from the very complex system of the 'real' world, and therefore cannot have an unique solution is often missed in discussion. This is generally true unless the 'real' system has a level of complexity equal to or below that of the model. This can happen where the observed data changes very slowly and smoothly - conditions which lead to a successful simulation using a 'metric' model. These models fit a curve to observed streamflow data, e.g. Littlewood 1993, and do not have explicit storage or loss functions. All models have the possibility of

an acceptable performance if the scale and definition of the output are coarse enough so that detail is lost.

During a discussion after the meeting Professor Bevan appeared to favour a solution to the problem which did not involve automatic optimization, but this does involve procedures which are computer intensive and wasteful of computing resources. This type of procedure such as the Generalised Likelihood Uncertainty Estimation Procedure (GLUE) should be possible using large modern computers with batteries of array processors; these are very expensive to run and because of the cost and time involved not often used except by organisations with large funds to devote to a project. However, the funds for the solution of practical hydrologic problems are such that answers have to be found in a relatively short time on a PC. The author was interested to find that several divisions of the National Rivers Authority were experimenting with applications of HYRRROM to particular problems. The use of computationally efficient models and optimization algorithms is for all practical purposes a financial and operational necessity so that the five objectives of this Thesis outlined in the Preface are still as valid now as when the research was begun.

10.2 The problem of global optimization.

The automatic determination of a global optimum in the presence of local optima, or even that the optimum reached is only 'local' is a problem that appears to have only one solution: the complete search of feasible n-dimensional parametric space.

The global optimum exists if the following applies:-

$$\begin{aligned} &f(x) \text{ where } x \in \text{the set } S \\ &\text{then there } \exists \text{ a point } x^* \in S \text{ where} \\ &f(x) \geq f(x^*) \end{aligned} \quad (10.1)$$

Whereas the local minimum, $f(x_{min})$, is in a neighbourhood N

$$\begin{aligned} &x_{min} \in N \text{ which is } \subset S \\ &\text{and } f(x) \geq f(x_{min}) \\ &\text{for all } x \in N \end{aligned} \quad (10.2)$$

However, if a local optimum satisfies the criterion for validation of the model given in equation 1.1

$$| \text{observed} - \text{simulated} | \leq \epsilon$$

then the objective of the simulation has been achieved. A further search to determine if there is a global optimum would appear to be a pointless exercise unless the parameters of the model require further refinement to enable a better understanding of the relationship between model concepts and the reality of the environment.

A paper by Dixon, Gomulka and Szego (1975) said that the determination of the global optimum was 'in its infancy', and other papers given at the same conference to estimate the global optimum by analysis either simply discussed the problem or arrived at solutions in very restricted cases.

The possible use of modern optimization techniques was mentioned at the British Hydrological meeting on 'Rainfall/runoff modelling' 1993. Examples given were genetic

algorithms, those using simulated annealing algorithms and neural networks to solve the problem of the global optimum, but without any further development of the theme. The use of the Rechenberg Evolution algorithm in this study must be one of the first examples of the use of a genetic algorithm in hydrology.

A paper on the use of a neural network package was presented by Hall and Minns (1993) as an application of artificial intelligence to rainfall/runoff modelling. A sweeping generalisation was made that the application compared favourably with results from conceptual models, but the two applications given were so limited that they could not be considered to substantiate this claim. The use of a 'loss function' such as the Penman open water evaporation estimate was not considered, and the systems were closed to any losses. The method appears to have little to commend it over the use of a multiple regression such as that by Wright (1975), and the latter method was probably simpler and not so demanding of computer resources. To understand the problems posed in using a network of synapses and neurons as the basis of 'artificial intelligence' a good study and discussion has been presented by Penrose (1989,1994). To claim for the neural network method that 'this is the way the brain works' is possibly very premature, but in the limited sense of pattern recognition it could well be true. Further research needs to be done before applying the method routinely in the field of hydrology.

However, these modern techniques seem to have little advantage, if any, over a straight forward grid search of the whole of feasible n-space in identifying the global optimum.

10.3 The limits and successes of HYRROM.

The ability of the nine parameter model, with six other fixed parameters, to simulate flows from such diverse catchments as the three test data sets, is exceptional. The data sets differ widely in location, altitude, vegetation, soil type and the demands of the evapotranspiration 'loss' functions based on the Penman index of evaporation from open water.

The model as set up was originally intended to be used to test various optimization algorithms for the purposes of this research. It was first mounted as a simple model package for use in flow forecasting on the island of St. Lucia, and was then found to be of use elsewhere and mounted as the interactive package HYRROM with the Rosenbrock algorithm to optimize the parameters. The limits of 'feasible' space to search for optimal values were fixed at ± 0.5 of the parameter values in the package shown in Table 9.1. It has been demonstrated during this research that these constraints needed relaxing during the course of the series of three tests, even if the final value was found in the original feasible space, and this facility should be included in the HYRROM package.

A notable failure with the full fifteen parameter model, HYRROM with all parameters 'floated', was the failure to model the Chyulu Hills volcanic aquifer in Kenya. This is over 2,000 Km² in extent, and mainly composed of volcanic clinker and ash so that the dominant process is vertical infiltration at the surface. It was later found, Eeles and Blackie (1988), that the model had failed because the full overall delay for the aquifer was of the order of 132 months with an initial peak for the 'rapid' response at 36 months. Both the large and small

models can only have maximum responses of a calendar month due to their only processing a month of data at a time.

10.4 The land use model.

This model was developed as a relatively simple means of investigating changes in land use and is basically a routing and storage shell for the Calder and Newson (1979) and Calder (1986) equations. The smaller parameter model is included in the model as a subset with parameters representing the area of grassland. The model was first used in the determination of the afforestation upper limit for the gathering grounds of the Elan Valley reservoirs, Eeles, Farquharson and Harding (1986), a commissioned project for the then Severn/Trent and Welsh Water Authorities. The simulated output for the period April 1932 to December 1984 from the model for seven scenarios for different levels of afforestation was then used, Eeles and Douglas (in preparation), to determine the level of operations and financial investment constraints for the two Authorities.

The model was based entirely on the parameters determined during the calibration period, April 1969 to December 1976, and its performance could only be assessed by using independent measures of the evapotranspiration loss. Reliable data sets with definite dated areas of clearance of vegetation are very scarce, and the model was tried on a biased set of flow data from the Hore in the Plynlimon experimental catchments on which it performed satisfactorily in predicting the effect of forest clearance on differences caused to low flows.

The model performed well on the Kirkton, Balquhiddy, data simulating the partial clearance of forest with a close

approximation to the change in flows. The soil moisture deficit readings obtained by neutron soil moisture meter, Eeles (1969), from experimental plots in the catchments are approximately the same as the model simulation values which represent the whole area of vegetation, and changes and trends in soil moisture deficit (SMD) are followed well.

The two models are interchangeable in the same modelling Library, but are not intended to do the same thing. The small nine parameter model is intended to give an acceptable simulation of the stream flows assuming a quasi-steady state in the catchment, whilst the large model allows some estimate of the effects of land use changes such as clearance (with assumptions on re-colonisation) and afforestation. The larger model is better, but not significantly so, on the Kirkton catchment. This is possibly an indication that the limit has been approached to which the observed data can be simulated due to its inherent errors and bias.

10.5 The sets of test data.

It is necessary to recognise the interdependence of the three elements of this type of modelling: the model, the data and the optimization algorithm. The objective function links the model through the simulated data and the observed flow data, and this in turn generates the function space which is searched by the optimization algorithm and so each element cannot be considered in isolation. Three forms of objective function were used in the land use model: the normal 'least squares' for the statistics and criteria for convergence, an 'eighth power' version for 'rapid' response parameters, and a proportional function used for parameters simulating the slowly

changing low flows. A logarithmic objective function was tested but gave unsatisfactory results because the changing sensitivity of the function could not be matched to fit that of the hydrograph. The objective function for HYRRROM remained as the 'least squares' version.

The use of the efficiency of fit, or coefficient of determination, changing by less than a preset value as a criterion for convergence, greatly reduced the number of model runs. It replaced the criterion for convergence initially built into the Rosenbrock algorithm of each parameter changing by less than a given value. This was found to generate a large number of wasted model runs which did not alter the objective function significantly or affect the fit of the model. However, this process was more logical and efficient than the original paper by Rosenbrock using the succession of ten 'failures' along each axis as the criterion for convergence.

A further great reduction in the time taken for each model run was the use of implied 'DO' loops to give the monthly input data as an unformatted block with a small lead block containing the month, year and number of days in the month to enable the large data block to be processed. The use of direct 'DO' loops generates computer instructions between each item of data instead of the two sets now used for each block.

Chapter 8 has been devoted to the testing of the data to ensure that it is homogenous and consistent over the period of the data set, and also to showing a way to extract fixed values from the observed data for the two parameters affecting the shape of the stream recession curves in dry periods.

The use of generated data sets with test 'errors' introduced was rejected in favour of the use of actual observed

data. This was done as problems with the algorithms were expected to emerge with 'real' data rather than with 'smooth' generated data. A good example of this was the Evolution algorithm which worked excellently with the Rosenbrock parabolic valley problem, but had to be substantially modified to work with the models and catchment data.

10.6 General problems with the optimization algorithms.

The development of two libraries to contain these algorithms has given a powerful tool for finding and testing an optimal solution to hydrologic problems. The libraries are independent of the model programs used since the two are connected by use of a large common block which can be varied to fit the model used.

The algorithms needed considerable work on their original program code to remove errors and get them running efficiently or indeed start to run! This is strange since they were all published program code in reasonably standard FORTRAN, but the move from large mainframes and different compilers to a PC and Microsoft FORTRAN 5.1 proved that they were not portable. Each algorithm had to be examined in detail to ensure that it was operating correctly in accord with the description given by its author or publisher, and that its array dimensions were correct. If arrays overflow then the program area of the computer can be overwritten, and this gives rise to errors which are beyond the diagnostics of compilers; different errors appear with any change to the code such as simple 'PRINT' commands. Optimal values of program parameters had to be found to fit the algorithms to hydrological problems rather than the problem for which they had first been designed. In some cases

names of algorithm parameters were taken from theory as input and given an entirely different computer mnemonic in the program. e.g. Input **H** became **DSTEP** in the version of the Marquardt program so that the latter became zero.

10.7 Results of the tests.

The first algorithm tested was the Rechenberg Evolution strategy, which on the basis of a test run on the Rosenbrock parabolic valley appeared to show considerable promise. The algorithm is so large that it needed its own library of subroutines.

The main problem with the published code was that it had lost any randomised determination of succeeding values of the random function seed, so that twenty succeeding calculations of the random function produced exactly the same output. This produces a statistical mean with zero deviation! The output from the random function needed scaling by a constant so that the full algorithm came into use otherwise only one of the complicated pathways through the algorithm was followed because of the small size of the random output and its effect on the choice of program branch. A final modification was the scaling of model parameters which achieved the best reduction of the objective function.

The performance of this algorithm was disappointing with its final ranking of sixth place. It did, however, produce a positive correlation in Figure 9.6 having started with negative values for the four and five parameter tests which gave some idea of its robustness. The use of randomized search vectors in a sector of the hyperellipsoid appears to be the weakness of

this algorithm. Perhaps a more regular search pattern in the sector would prove more successful.

The other library of five modified algorithms demonstrated the superiority in the tests of the Rosenbrock, Direct Search, and Simplex over the slope dependent Conjugate Gradient and Marquardt algorithms. The initial data errors probably caused ill-conditioning in the slope elements:-

$$\nabla F(X_j) = \frac{\partial(F_j + e_j)}{\partial x_j} \quad 1 \leq j \leq n \quad (10.4)$$

where n is the number of active parameters, and e_j is the initial data error.

This error in the objective function caused by initial data errors is positive, as they are all even power functions, and therefore leads to error in the slope of steepest descent calculated from the data. The 'ripple' in the three dimensional plot of the two delay parameters shown in Figure 5.2 shows the problem posed by the topography of the n -dimensional space in determining slopes.

The three pattern search algorithms each have a definite search strategy which does not require differentials and is based purely on the 'success' or 'failure' of the trial search vector in finding a new lower value of the objective function.

10.8 Hybridization of algorithms.

The values of the Rosenbrock algorithm for the nine parameter tests were used as initial data for the algorithms hybridized with each of the other algorithms except the Evolutionary strategy. The large number of model runs required for this strategy would appear to put it out of contention as a partner in a hybridized algorithm. The Simplex algorithm appeared from these tests the most promising in reduction of the objective function value and volume error. However, this was at the expense of a large number of model runs.

The Simplex was hybridized with the Rosenbrock algorithm and a full series of tests run on the three sets of data. This succeeded in producing significant reductions in the objective function and volume error but put up the 'overhead' for model runs by some 70% for the total of the tests.

10.9 Recommendations based on the tests.

It is impossible to say that the findings of this research are universally valid because of the interdependence of models, observed data and optimization algorithms. However, the data sets chosen vary considerably in vegetation, geographical location and altitude, input/output and system losses so that the work should at least provide a strong indication of how to approach hydrologic simulation problems.

For this large data set type of hydrologic problem the recommended algorithm parameters and modifications should be included. It is unfortunately true that an algorithm will work if it is capable of retaining the lowest value of objective function amongst those found, but it is not necessarily following the strategy that has been outlined.

Published code should not be mounted and used without critical appraisal, and for the same reason the output from already compiled libraries should be treated with caution. The algorithms are almost certain to have been 'fitted' to particular problems which are very different to those caused by hydrologic data!

The models have shown their usefulness and accuracy together with their low use of computing resources in providing simulations on which pragmatic decisions can be based. In the absence of large project funding they are indeed the easiest and reliable way forward provided that there is enough length of reliable data.

10.10 Recommendations for future work.

Unfortunately, this research has proved to be rather like painting the Forth Bridge - it is never complete! One of the problems is that with each new finding previous work has had to be repeated to see if it is still valid. It is surprising that with the demand for practical solutions to water resource problems so little research has been done on the interrelationship of the three elements of model, data and optimization routine.

The slope dependent algorithms need investigating in the context of large observed data sets with significant initial error to confirm if this is the reason for the ill-conditioning of their rapid solutions.

The effects of scaling the model parameters for the Simplex and Conjugate Directions algorithms should be tested. The Simplex algorithm works with limits and without scaling, whilst the other works without either, and is completely

unconstrained. Is a significant difference made by including scaling and constraints?

The Evolutionary strategy, like the Theory, requires the determination of a better focus for its search than random selection, but there are so many parameters built into this code that further work is probably not worthwhile.

The model concepts are coarse but fit the scale of the problem. Work should continue on determining the range of various model parameters rather than seeking a universal value for each one. Their concepts for integrated areas could be used in conjunction with the large 'physically' based models on areas where there is little change in the field parameters required for these models.

The Rosenbrock algorithm has long been discarded for other types of problem, but seems to be ideally suited for hydrologic modelling. There must be limits to the use of this algorithm for large data sets and these should be further investigated.

This discipline is essentially pragmatic - answers have to be found to hydrologic problems, and decisions taken on that basis and implemented at least financial cost. How much one can rely on those solutions can only be tested by the 'split record': a short period of data over at least a 'water year' is used to calibrate the model, and a longer period for validation. Again this is a practical solution to a very complicated stochastic problem.

REFERENCES

(a). Published work and reports by C.W.O.Eeles.

Eeles, C.W.O., 1969. "Installation of access tubes and calibration of Neutron Moisture Probes", Report No. 7, Institute of Hydrology.

Eeles, C.W.O., 1978. "A conceptual model for the estimation of historic flows", Institute of Hydrology report No. 55.

Eeles, C.W.O., 1979. "Soil moisture". Paper in E.Afr.Agric.For.J. Vol. 43 (Special issue).

Eeles, C.W.O., 1983. "A conceptual model for the river Thames catchment to Eynsham". Commissioned report to the Department of the Environment.

Eeles, C.W.O., F.A.K.Farquharson, R.J.Harding, 1986. "Report on the effects of progressive afforestation on runoff from the Elan Valley Catchments, commissioned report to the Severn/Trent and Welsh Water Authorities.

Eeles, C.W.O. and Y.Parks, 1986. "The St. Lucia rainfall/runoff model". Commissioned report for the Overseas Development Administration, UK Foreign Office, 35 pp.

Eeles, C.W.O., E.Munir, J.R.Blackie, 1988. "The development of a rainfall model". Joint commissioned report with BGS to ODA "The Chyulu Hills Water Resources Study, Kenya: 1984-87", Vol.II, Chapter 7, pp 21.

Eeles, C.W.O., and J.R.Blackie, 1988. "The development of a recharge model". Ibid. Chapter 9, pp 22.

Eeles, C.W.O., and J.R.Blackie, 1988. "The correlation of model simulated recharge with spring flows". Ibid. Chapter 11, pp.16.

Eeles, C.W.O., Y.Parks and A.Barr, 1989. "HYRROM Operation Manual". Institute of Hydrology, Wallingford, Oxfordshire.

Eeles, C.W.O., M.Robinson and R.C.Ward, 1990. "Experimental Basins and Environmental Models". Keynote lecture paper in 'Hydrological Research Basins and the Environment', Proceedings of the International Conference, 24-28 September 1990, International Hydrological Programme of UNESCO, Proceedings of the TNO Committee on Hydrological Research, Wageningen, Netherlands, No. 44, pp 3-12.

Eeles, C.W.O., 1990. "Simulation of catchment runoff by conceptual model". In 'Computer modelling in the environmental sciences', D.G.Farmer and M.J.Rycroft, editors, IMA Conference Series, Oxford University Press, p.103-109.

Eeles, C.W.O. and J.R.Douglas 1992. "The modelling of afforestation, and its effect on the water resources of the Elan Valley Reservoirs". (In preparation).

Eeles, C.W.O. and J.R.Blackie, 1993. "Land use changes in the Balquhiddy catchments simulated by a daily streamflow model". Journal of Hydrology, Vol. 145, Issue 3-4, pp. 315-336.

Joint author with:-

Blackie, J.R. and C.W.O.Eeles, 1985. "Lumped catchment models". Ch. 11 of 'Hydrological Forecasting', Editors M.G.Anderson and T.P.Burt, John Wiley and Sons Ltd., pp 311-345.

Blackie, J.R., C.W.O.Eeles, A.M.Roberts, A.Matthews, T.K.M.Simpson, C.Hughes, and K.B.Burton, 1990. "Experimental Catchments Yearbook 1986". "Hydrological Review", Section 2 by C.W.O.Eeles. Published by Institute of Hydrology, Wallingford.

Edwards, K.A., J.R.Blackie, and C.W.O.EELES, 1978. "Final report on the East African catchment research project (ODM R2582)". IH Report to ODM 4 vol, 625 pp. Subsequently published as : "Hydrological research in East Africa", E.Afr. Agric.For.J., Vol. 43 (Special Issue), Muguga, Nairobi.

Gross, R., C.W.O.Eeles, and A.Gustard 1989. "Application of a lumped conceptual model to FRENED catchments". IASH Publication 187: pp. 309-320.

Gustard, A., L.A.Roald, S.Demuth, H.S.Lumadjeng and R.Gross 1989. "Modelling the impact of afforestation and clear felling". Chapter 7.3.5 Contributed by C.W.O.Eeles to 'Flow Regimes from Experimental and Network Data' (FRENED), Vol 1: pp 244- 270.

Robinson, M., C.Eeles and R.C.Ward 1990. "The research basin and stationarity". In 'Hydrological Research Basins and the Environment', Proceedings of the International Conference, 24-28 September 1990, International Hydrological Programme of UNESCO, TNO Committee on Hydrological Research, Wageningen, Netherlands.

(b). General references.

Adby, P.R. and M.A.H.Dempster, 1974. "Introduction to Optimization Methods". Chapman and Hall, London.

Abbott, M.B., J.C.Bathurst, J.A.Cunge, P.E.O'Connell and J.Rasmussen, 1986. "An introduction to the European Hydrological System". J.Hydrol. 87: 45-49.

Ahrens, J.H., and U.Dieter, 1972. "Computer methods for sampling from the exponential and normal distributions". CACM 15, p. 873-882 and 1047.

Beven, K. 1989. "Changing ideas in hydrology - The case of physically based models". J.Hydrol. 105: 157-172.

Blackie, J.R. 1972. "The application of a Conceptual Model to two East African Catchments". An unpublished report submitted in partial fulfillment of the requirements for M.Sc. of the University of London.

Blackie, J.R. and C.W.O.Eeles, 1985. "Lumped catchment models". Ch. 11 of 'Hydrological Forecasting', Editors M.G.Anderson and T.P.Burt, John Wiley and Sons Ltd., pp 311-345.

Blackie, J.R., C.W.O.Eeles, A.M.Roberts, A.Matthews, T.K.M.Simpson, C.Hughes, and K.B.Burton, 1990. "Experimental Catchments Yearbook 1986". "Hydrological Review", Section 2 by C.W.O.Eeles. Published by Institute of Hydrology, Wallingford.

Blackie, J.R., 1993. "The Water Balance of the Balquhiddy catchments", Journal of Hydrology, 145 (1993), pp 239-257, Elsevier.

Bunday, B.D., 1984. "Basic optimisation methods". Edward Arnold (Publishers) Ltd.

Broyden, C.G., 1969. "Quasi-Newton methods and their application to function minimization". Math. Computing, 21, p 368-381.

Boughton, W.C., 1965. "A new simulation technique for estimating catchment yield". University of New South Wales Water Research Lab. Report 78.

Box, M.J., D.Davies and W.H.Swann, 1969. "Nonlinear optimization techniques", ICI Monograph 5, published by Oliver and Boyd, Edinburgh.

Calder, R.J.Harding and P.T.W.Rosier, 1983. "An objective assessment of soil moisture deficit models". J.Hydrol., 60, pp 329-355.

Calder, I.R. and M.D.Newson 1979. "Land use and upland water resources in Britain - a strategic look". Water Resources Bulletin, 16: pp 1628-1639.

Calder, I.R., 1986. "The influence of land use on water yield in upland areas of the U.K." J.Hydrol., 30, pp 201-211.

Calver, A. and W.L.Wood, 1989. "On the discretization and cost- effectiveness of a finite element solution for hillslope subsurface flow". J.Hydrology, 110: pp. 165-179.

Calver, A. and P.Binning, 1990. "On modelling hillslope water flow paths and travel times". J.Hydrology, 121: pp 335-344.

Clarke, R.T. 1973. "Mathematical models in hydrology". Irrigation and drainage paper No. 19, Food and Agriculture Organisation of the United Nations, Rome.

Dixon, L.C.W., Gomulka, J., and Szegö, G.P, 1975. "Towards global optimisation", in the book of the same name edited by L.C.W.Dixon and G.P.Scegö, North-Holland, pp 29-54.

Edwards, K.A., J.R.Blackie, and C.W.O.EELES, 1978. "Final report on the East African catchment research project (ODM R2582)". IH Report to ODM 4 vol, 625 pp. Subsequently published as : "Hydrological research in East Africa", E.Afr. Agric.For.J., Vol. 43 (Special Issue), Muguga, Nairobi.

Darcy, H., 1856. "Les Fontaines Publiques de la Ville de Dijon", Dalmont, Paris.

Dawdy, D.R. and T.O'Donnell, 1965. "Mathematical models of catchment behaviour". J. Hyd. Div. Proc. ASCE, 91, HY4, 123.

- Dennis, jr, J.E., and R.B.Schnabel, 1983. "Numerical methods for unconstrained optimization and nonlinear equations". Prentice-Hall Series in Computational Mathematics.
- Dickinson, W.T. and J.R.Douglas, 1972. "A conceptual runoff model for the Cam catchment". Inst. Hydrol. Report No. 17.
- Dooge, J.C.I., 1959. "A general theory of the unit hydrograph". J. Geophysical Res., 64(1), 241.
- Dooge, J.C.I., 1973. "Linear Theory of Hydrologic Systems". Tech. Bull. No. 1468, Agricultural Research Service, United States Dept. of Agriculture.
- Dunin, F.X., 1975. "Use of Physical Process Models" in 'Prediction in Catchment Hydrology', Ed. T.G.Chapman and F.X. Dunin, Australian Academy of Science, pp 277-291.
- Farmer, D.G. and M.J.Rycroft, 1990. "Computer modelling in the environmental sciences", D.G.Farmer and M.J.Rycroft, editors, IMA Conference Series, Oxford University Press.
- Fleming, G., 1975. "Computer simulation techniques in Hydrology", Elsevier Publishing Co. Amsterdam.
- Fletcher, R., 1987. "Practical methods of optimization", 2nd. Edition, John Wiley and Sons.
- Gill, P.E., W.Murray and M.H.Wright, 1981. "Practical optimization". Academic Press Inc. (London) Ltd.
- Gill, P.E., W.Murray, M.H.Wright, 1991. "Numerical linear algebra and optimization". Vol. 1, Addison-Wesley Publishing Company.
- Gross, R., C.W.O.Eeles, and A.Gustard 1989. "Application of a lumped conceptual model to FRENDE catchments". IASH Publication 187: pp. 309-320.
- Gustard, A., L.A.Roald, S.Demuth, H.S.Lumadjeng and R.Gross 1989. "Modelling the impact of afforestation and clear

felling". Chapter 7.3.5 Contributed by C.W.O.Eeles to 'Flow Regimes from Experimental and Network Data' (FRIEND), Vol 1: pp 244- 270.

Gustard, A. and A.J.Wesselink, 1993. "Impact of land use on water resources: Balquhiddar". Journ. Hydrol. 145 (1993), pp 389-401, Elsevier.

Hall, M.J. and A.W.Minns, 1993. "Rainfall-runoff modelling as a problem in artificial intelligence: experience with a neural network". BHS 4th. National Hydrology Symposium, Cardiff.

Hall, R.L., 1987. "Processes of evaporation from vegetation of the uplands of Scotland". Trans. Roy. Soc. Edinburgh, Earth Sciences, 78, pp 327-324.

Hall, R.L. and R.J.Harding, 1993. "The water use of the Balquhiddar catchments: a processes approach". Journal of Hydrology, 145, (1993), pp 285-314, Elsevier.

Herodotus of Halicarnassus, 445 BC. "Histories", Translated by Wesseling, fol. Amsterdam 1793.

Hillel, D., 1977. "Computer simulation of soil water dynamics; a compendium of recent work". International Development Research Centre, Ottawa.

Horton, R.E., 1933. "The role of infiltration in the hydrologic cycle". Trans. Am. Geophys. Union, 14, 446-460.

Ibbitt, R.P., 1972. "Effects of random data errors on the parameter values for a conceptual model". Water Resources Res., 8, 70-78.

Ibbitt, R.P., and T.O'Donnell, 1971. "Fitting methods for conceptual catchment models". J. Hydraul. Div. Am. Soc. Civ. Engrs., 97, 1331-1342.

Ibbitt, R.P., and T.O'Donnell, 1971. "Designing conceptual catchment models for automatic fitting methods". IASH Publ., 101, 461-475.

Johnston, P.R. and D.H.Pilgrim, 1973. "A study of parameter optimisation for a rainfall runoff model". Report No. 131, Australian Water Resources Council Research Project 68/1.

Johnson, R.C., J.R.Blackie, and J.A.Hudson, 1990. "Methods of estimating precipitation inputs to Balquhider experimental basins, Scotland". 'Hydrology in mountainous regions' Vol. 1. Hydrological measurements: the water cycle. Editors H.Lang and A.Musy, IASH publications No. 193, pp 7-14.

Johanson, R.C., J.C.Imhoff and H.H.Davis Jr, 1980. "User's manual for the Hydrological Simulation Program - Fortran (HSPF)". Environmental Protection Agency Report FPA-600/9-80.

Johnson, W.J. and R. Dean Riess, 1982. "Numerical Analysis", Addison-Wesley Publishing Company.

Johnson, W.J., R.D.Riess and J.T.Arnold, 1989. "Introduction to Linear Algebra". Addison-Wesley Publishing Company.

Jones, D.A., 1983. "Statistical analysis of empirical models fitted by optimization". Biometrika 70, 1, pp 1-22.

Kendall, M.G., and P.A.P.Moran, 1963. "Geometrical Probability". Vol. 10 Griffin's Statistical Monographs & Courses, Ed. M.G.Kendall for Charles Griffin & Co. Ltd.

Kirby, C., M.D.Newson, and K.Gilman, 1991. "Plynlimon Research: the first two decades". Institute of Hydrology Report No. 109.

Kuczera, G., 1987. "Fitting and testing mathematical models: a user manual for program suite NLFIT". Version 2, Department of Civil Engineering and Surveying, The University of Newcastle, N.S.W., 2308, Australia.

- Kuczera, G., 1990. "Assessing hydrologic model nonlinearity using response surface plots". J. Hydrol., 118, 143-161.
- Leonardo da Vinci, 1508-1519. Notebooks of Leonardo da Vinci, not published until the 19th. Century. Dibner, B., 1946. "Leonardo da Vinci, Military Engineer", New York: The Burndy Library.
- Linsley, R.K. and N.H.Crawford, 1960. "Computation of a synthetic streamflow record on a digital computer". Hydrol. Sci. Bull. IASH, Pub. 51, pp 526-538.
- Crawford, N.H. and R.K.Linsley, 1962. "The synthesis of continuous streamflow hydrographs on a digital computer". Dept. of Civil Engineering, Stanford University, California, Technical Report No. 12.
- Crawford, N.H. and R.K.Linsley, 1964. "A conceptual model of the hydrologic cycle". Hydrol. Sci. Bull, IASH, Pub. 63, pp 573-587.
- Crawford, N.H. and R.K.Linsley, 1966. "Digital Simulation in Hydrology", Stanford Watershed Model IV, Dept. of Civil Eng., Stanford University, Technical Report No. 39.
- Mandeville, A.N., P.E.O'Connell, J.V.Sutcliffe and J.E.Nash, 1970. "River flow forecasting through conceptual models. Pt. III - The Ray catchment at Grendon Underwood". J. Hydrol., 11, 109-128.
- McCarthy, G.T., 1938. "The unit hydrograph and flood routing". Proc. Conf. North Atlantic Division, U.S. Army Corps of Engineers.
- McCulloch, J.S.G., 1965. "Tables for the rapid computation of the Penman estimate of evaporation". East African Agriculture and Forestry Journal, Vol. XXX, pp 286-295.

- Morris, E.M., 1983. "Modelling the flow of mass and energy within a snowpack for hydrological forecasting". *Ann. Glaciol.*, 4, 193-203.
- Morris, E.M., 1985. "Snow and ice". Ch. 7 of 'Hydrological Forecasting', Editors M.G.Anderson and T.P.Burt, John Wiley and Sons Ltd., pp 153-182.
- Mulvaney, T.J., 1851. "On the use of self registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment." *Proc. Inst. Civil Engineers, Ireland*, 4, pp 18-31.
- Nash, J.E., 1957. "The form of instantaneous unit hydrograph". *Hydrol. Sci. Bull.*, 3, pp 114-121.
- Nash, J.E. and J.V.Sutcliffe, 1970. "River flow forecasting through conceptual models. Pt. 1 - A discussion of principles". *J.Hydrol.* 10, 282-290.
- Nelder, J.A. and R.Mead 1965. "A simplex method for function minimisation". *Computer J.*, 7: pp 308-313.
- O'Connell, P.E., A.N.Mandeville, J.V.Sutcliffe and J.E.Nash, 1970. "River flow forecasting through conceptual models. Pt. IV - The Brosna catchment". *J.Hydrol.* 11, 129- .
- O'Connell, P.E., 1980. "Real time hydrological forecasting and control". *Proc. Int. Workshop Inst. Hydrology*.
- Parkinson, J.M. and D.Hutchinson, 1972. "An investigation into the efficiency of variants of the Simplex Method" in 'Numerical methods for non-linear optimization', F.A.Lootsma, Editor, Academic Press, London, pp 115-135..
- Penman, H.L., 1948. "Natural evaporation from open water, bare soil and grass". *Proc. Roy. Soc. A*, 193, 120-146.

Penman, H.L., 1949. "The dependence of transpiration on weather and soil conditions". J.Soil Science, Oxford, Vol 1, pp 74-89.

Penrose, R., 1989. "The Emperor's New Mind". Oxford University Press.

Penrose, R., 1994. "Shadows of the Mind". Oxford University Press.

Phillip, J.R., 1954. "An infiltration equation with physical significance". Soil Sci., 77, 153-157.

Pitman, W.V., 1973. "A mathematical model for generating monthly river flows from meteorological data in South Africa". Hydrol. Unit, Univ. Witwatersrand, Johannesburg, Rep. No. 2/73.

Pitman, W.V., 1976. "A mathematical model for generating daily river flows from meteorological data in South Africa". Hydrol. Unit, Univ. Witwatersrand, Johannesburg, Rep. No. 2/76.

Pitman, W.V., 1978. "Flow generation by catchment models of differing complexity and comparison of performance". J.Hydrol., 38: pp 59-70.

Pickup, G., 1977. "Testing the efficiency of algorithms and strategies for automatic calibration of rainfall/runoff models". Hydrol. Sci Bull., 22/2.

Powell, M.J.D., 1964. "An efficient method for finding the minimum of a function of several variables without calculating derivatives". Computer Journal, Vol. 7, 2, p 155-162.

Powell, M.J.D., 1965. "A method for minimizing a sum of squares of nonlinear functions without calculating derivatives". Computer Journal, Vol. 7, p 303-307.

Powell, M.J.D., 1969. VA05a, Harwell sub-routine library, release 10.

Rechenberg, I., 1964. "Cybernetic solution path of an experimental problem". Translated 1965 by B.F.Toms, Library translation No. 1122, RAE Farnborough.

Rechenberg, I., 1973. "Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution", Frommann-Holzboog, Stuttgart.

Refsgaard, J.C., S.M.Seth, J.C.Bathurst, M.Erlich, B.Storm, G.H.Jørgensen, and S.Chandra, 1992. "Applications of SHE to catchments in India". Part I, General Results. J.Hydrol. 140, pp 1-23.

Reklaitis, G.V., A.Ravindran, and K.M.Ragsdell, 1983. "Engineering Optimization", John Wiley and Sons.

Richards, L.A., 1931. "Capillary conduction of liquids through porous mediums". Physics, 1, 318-333.

Richards, L.A., W.R.Gardner, and G.Ogata, 1956. "Physical processes determining water losses from soil". Soil Sci. Soc. Am. Proc., 20, 310-314.

Rijtema, P.E., 1965. "An analysis of actual evapotranspiration". Agricultural report No. 659, Wageningen, Netherlands.

Roberts, G., M.France, R.C.Johnson, and J.T.Law, 1993. "The analysis of remotely sensed images of the Balquhiddy catchments for estimation percentages of land cover types". Journal of Hydrology, 145, 1993, pp 259-265, Elsevier.

Robinson, M., C.Eeles and R.C.Ward 1990. "The research basin and stationarity". In 'Hydrological Research Basins and the Environment', Proceedings of the International Conference, 24-28 September 1990, International Hydrological Programme of UNESCO, TNO Committee on Hydrological Research, Wageningen, Netherlands.

Rosenbrock, H.H., 1960. "An automatic method of finding the greatest or least value of a function". Computer Journal, 3, p 175-184.

Saint-Venant, B. de, 1871. "Theory of unsteady water flow, with applications to river floods and to propagation of tides in river channels." Comptes Rendus Acad.Sci., Paris, 73, pp 148-154, pp 237-240.

Schwefel, H.P., 1981. "Numerical optimization of computer models". John Wiley and Sons, Ltd.

Sherman, L.K., 1932. "Streamflow from rainfall by a unit hydrograph method". Eng. News Record, 108, pp 501-505.

Son Il, 1990. "Modelling the hydrological effects of land use change in small catchments". Unpublished PhD. thesis University of Southampton.

Sugawara, M., 1961. "An analysis of runoff structure about several Japanese rivers". Japanese J. Geophys., 2.

Swann, W.H., 1964. "Report on the development of a new direct searching method of optimization". ICI Ltd., Central Instrument Laboratory Research Note 64/63, Middlesborough, Yorks.

Thom, A.S., and Oliver, H.R., 1977. "On Penman's equation for estimating regional evaporation". Quart. J. R. Met. Soc. (1977), 103, pp 345-357.

Thorntwaite, C.W., 1948. "An approach towards a natural classification of climate". Geogr. Rev., 38, 55.

Wheater, H., 1993. 'Circulation', introduction to report on BHS national meeting on 'Rainfall/runoff modelling', R.Austin & T.Jolley pp. 15-20.

World Meteorological Office, No. 429, 1975. "Intercomparison of conceptual models used in operational Hydrological Forecasting". Office of Hydrological Research No. 7, Geneva.

World Meteorological Office, No. 646, 1986. "Intercomparison of models of snowmelt runoff". Office of Hydrological Research, No. 23.

World Meteorological Office, No. 740, 1990. "Hydrological models for water resources system design and operation". Office of Hydrological Research, No. 34.

World Meteorological Office, No. 779, 1987. "Simulated real-time intercomparison of hydrological models". Office of Hydrological Research, No. 38.

Wright, C.E., 1975. "The assessment of regional groundwater schemes by river flow regression equations". J.Hydrol. 22, pp 205-215.

Zeigler, B.P., 1976. "Theory of Modelling and Simulation", John Wiley and Sons.

Appendix 1.

The Penman equation.

The Penman equation, Penman (1948), was the first physically related treatment of the difficult problem of estimating evaporation from natural surfaces, and although semi-empirical the equation was determined from meteorological data. This equation expresses evaporation rate from an open water surface as a function of the net flux of radiant energy at the surface and an aerodynamic term representing the ventilation of the surface by the air in motion over it.

The basic equation elements are

E_o rate of evaporation from an open water surface

E_{aer} aerodynamic or ventilation term

E_{rad} evaporation equivalent of the net flux of radiant energy to the surface

e actual vapour pressure 2m above the surface

e_s saturation values of vapour pressure 2m above the surface

U wind run in km per day

Δ slope of saturation vapour pressure versus temperature curve for water at air temperature

γ thermodynamic value of the psychometric constant

The rate of evaporation from an open water is then

$$E_o = \frac{(E_{rad} \cdot \Delta + \gamma \cdot E_{aer})}{(\Delta + \gamma)} \quad (1)$$

and the corresponding aerodynamic term is

$$E_{aer} = 0.2625 (e_s - e) \left(1 + \frac{U}{160} \cdot 9\right) \quad (2)$$

Penman developed the expression from this for potential evapotranspiration, E_T from short, green vegetation completely covering the ground and with a sufficient supply of water from the soil for its needs as

$$E_T = K.E_o \quad (3)$$

Where K is a factor which varies seasonally being high in summer and low in winter, approximately 0.8 and 0.6 respectively, and decreasing under drought conditions when the plant is under moisture stress.

Appendix 2.

```

C Version of HYRROM used in this research - see note at end of code.
*****
C
C Filename : MODEL EVOL                      Date : 03/11/88
C Author : C.W.O.EELES

C Description : MODEL USED BY OPTIMIZATION LIBRARY

C Libraries needed : ROS.LIB or EVOLS.LIB
C
C I/O stream numbers : 5, Screen INPUT, 6, Screen OUTPUT
C                      4, Data file, 8,OUTPUT,
C
C Control Parameters : INPUT.DAT, read into system in control program

C Subroutines called : READER, STAT
C
C
C
C*****

      SUBROUTINE MODEL
C
C          CATCHMENTS MODEL
C
C          INCLUDE "COMMON.DAT"
C          NUM=N
C          IF (IFIN.EQ.0) GOTO 5
C          IF (NUM.EQ.0) GOTO 12
C          IF (IFIN.EQ.1) GOTO 12
C
C          5 DO 11 J=1,NN
C              IF (KK(J).EQ.0) GOTO 11
C              K = KK(J)
C              YI(J) = Y(K)
C          11 CONTINUE
C
C          12 SS = YI(1)
C              RC = YI(2)
C              RDEL = YI(3)
C              RX = YI(4)
C              RK = YI(5)
C              FC = YI(6)
C              GSU = YI(7)
C              GSP = YI(8)
C              GDEL = YI(9)
C              RSTORE = 0.0
C              RVOL = 0.0
C              DC = 10.0
C              NRDEL = INT(RDEL)
C              FRDEL = RDEL - FLOAT(NRDEL)
C              NGDEL = INT(GDEL)

```

```

FGDEL = GDEL - FLOAT(NGDEL)
ND = NGDEL + 2
III = III + 1
IF (NRDEL.GT.0) THEN
  DO 14 I=1,NRDEL
14   SRVOL(I) = 0.0
  ENDIF
  FP      = 0.0
  FO      = 0.0
  CS=0.0
  IM      = 0
  IM1     = 0
  IMM     = 0
  ITD     = 0
  SSQ     = 0.0
  SUM     = 0.0
  SUMP    = 0.0
  ERSUMP  = 0.0
  TRAIN   = 0.0
  SEVAP   = 0.0
  SEPREDE = 0.0
  NNDAAYS = 0
  KKD     = 0

C
C
C   Start of model run
18  IM = IM + 1
    DO 20 I=1,31
    EVAP(I)=0.0
20   PRUN(I) = 0.0
    CALL READER
    IIR = IIR + 1
    IF (IM.LE.1) THEN

C       Set initial ground water store & output values.

    DO 24 J=1,ND
24   SGVOL(J) = RUNOFF(J)
    IF (NGDEL.LE.0) THEN
      RGVOL = SGVOL(1)+(1.0-FGDEL)*(SGVOL(1)-SGVOL(2))
    ELSE
      RGVOL = (1.0-FGDEL)*SGVOL(NGDEL)+FGDEL*SGVOL(NGDEL+1)
    ENDIF
    IF (GSP.NE.0.0) THEN
      GS = ((RGVOL**(1.0/GSP))*GSU) - RGVOL
    ELSE
      GS = GSU-RGVOL
    ENDIF
  ENDIF
  LL = 1
  LM = 1
  SDIRO = 0.0
  SUMM  = 0.0
  SUMPM = 0.0
  SUMEP = 0.0
  SUMEM = 0.0
  SRRAIN = 0.0

```

```

C Start of main loop of model. Calculations for one month overlay.
C
C   DO 78 IK = 1,ID
C
C   -----
C
C   MODEL SECTION 1.
C
C       (Interception)/ Surface storeage component.
C
C   'ERAIN' is rain passing beyond this store.
C   'SS' is the capacity of the store.
C   'CS' is the current content of the store.
C
C   -----
C
C       ERAIN = RAIN(IK)-SS+CS
C       IF (ERAIN.LE.0.0) ERAIN = 0.0
C       CS = RAIN(IK)-ERAIN+CS
C       ES=(1.0+FC)*EVAP(IK)
C       IF(ES.GE.CS) THEN
C           ES = CS
C       EEVAP=EVAP(IK)-CS/(1.0+FC)
C       ELSE
C           EEVAP=0.0
C       ENDIF
C       CS=CS-ES
C       ERAIN=RAIN(IK)+CS-SS
C       IF(ERAIN.LE.0.0) ERAIN=0.0
C       CS=CS+RAIN(IK)-ERAIN
C
C   -----
C
C   MODEL SECTION 2.
C
C       Calculation of actual Evap from EO+DC
C   'FC' is the factor reducing EO to potential transpiration.
C
C   -----
C
C   SOIL MOISTURE STORE
C
C       STORE=9.5
C       IF(EEVAP.GT.0.0) THEN
C           IF(CSTOR.GT.0.0) THEN
C               ECC=FC*EEVAP
C               IF(ECC-STORE.GE.0.0) THEN
C                   ECC=CSTOR
C               EEVAP=EEVAP-ECC/FC
C               CSTOR=0.0
C           ELSE
C               CSTOR=CSTOR-ECC
C           EEVAP=0.0
C       ENDIF
C       ECC=0.0
C       ENDIF
C       TNG=DC/350.0
C       IF(TNG.LT.0.0) TNG=0.0
C       IF(TNG.GT.3.1416) TNG=3.1416
C       ECP=FC*((COS(TNG)-1.0)/2.0)
C       EC = ECP*EEVAP

```

```

DC = DC + EC
ELSE
ECC=0.0
EC=0.0
ENDIF
EPRED(IK) = EC + ES+ECC
DEF(IK) = DC

```

MODEL SECTION 3.

Calculation of direct runoff from ERAIN

'DC' is mean catchment deficit in mm
 'RC' is max rop (at DC = 0.0)
 Routing of direct runoff to produce hydrograph.
 'RSTORE' is resevoir
 'RK' is its constant
 'RX' is its exponent
 'RDEL' is delay (in model intervals)
 RS & RR are exp. factors relating ROP & ERAIN

```

ROFF = 0.0
IF (ERAIN.GT.0) THEN
CSTOR=CSTOR+ERAIN
IF(CSTOR.LE.9.5) THEN
ROFF=0.0
ERAIN=0.0
ELSE
ERAIN=CSTOR-9.5
CSTOR=9.5
ENDIF
  ROFF = RC*ERAIN
  RSTORE = RSTORE + ROFF
ENDIF
IF (RSTORE.EQ.0.0) THEN
  RO = 0.0
ELSE
IF (RX.EQ.0.0) THEN
  RO = RK
ELSE
  RO = RK*RSTORE**RX
ENDIF
ENDIF
IF (RO.GT.RSTORE) RO = RSTORE
RSTORE = RSTORE - RO
IF (IK.LE.NRDEL) PRUN(IK) = SRVOL(IK) + PRUN(IK)
VOL = FRDEL*RVOL+(1.0-FRDEL)*RO
RVOL = RO
IF ((IK+NRDEL).GT.ID) THEN
  SRVOL(LM) = VOL
  LM = LM + 1
ELSE
  PRUN(IK+NRDEL) = PRUN(IK+NRDEL)+VOL
ENDIF
DIRO(IK) = ROFF
SDIRO = SDIRO+ROFF

```

MODEL SECTION 4.

Contribution to runoff from groundwater.

'GPR' is percolation rate to groundwater.

'GRO' is runoff from groundwater.

'GS' is current level of groundwater storage.

GRO is non-linear function of GS

'GSU' & 'GSP' are parameters in non-linear flow curve.

DC = DC-ERAIN+ROFF

IF (DC.GE.0.0) THEN

GPR = 0.0

ELSE

GPR = -DC

ENDIF

GRP=(DC+15.0)*0.1

IF(DC.LT.0.0) DCC=-DC

IF(DC.GE.0.0) DCC=1.0

GPR=0.4*(COS(GRP)+1.0)*DCC

IF(DC.LT.-23.0) GPR=15.7

IF(DC.GT.11.0) GPR=0.0

DC = DC + GPR

IF (GS.LE.0.0) THEN

GS = 0.0

GSTORE(IK) = 0.0

GRO = 0.0

ELSE

GSTORE(IK) = GS

IF (GSU.NE.0.0) THEN

GRO = (GS/GSU)**GSP

ELSE

GRO = 0.0

ENDIF

ENDIF

IF (IK.LE.NGDEL) PRUN(IK) = PRUN(IK)+SGVOL(IK)

GVOL = FGDEL*RGVOL+(1.0-FGDEL)*GRO

RGVOL = GRO

IF ((IK+NGDEL).GT.ID) THEN

SGVOL(LL) = GVOL

LL = LL + 1

ELSE

PRUN(IK+NGDEL) = PRUN(IK+NGDEL)+GVOL

ENDIF

GS = GS-GRO+GPR

SUMM = SUMM+RUNOFF(IK)

SUMPM = SUMPM+PRUN(IK)

SUMEP = SUMEP+EPRED(IK)

SUMEM = SUMEM+EVAP(IK)

SRRAIN = SRRAIN+RAIN(IK)

ERR(IK) = PRUN(IK)-RUNOFF(IK)

IF (RUNOFF(IK).EQ.0.0) THEN

ERPC(IK) = 1000.0

ELSE

```

      ERPC(IK) = 100.0*ERR(IK)/RUNOFF(IK)
    ENDIF
    IF (ERPC(IK).GT.1000.0) ERPC(IK) = 1000.0
    IF (IFIN.NE.0) ERSUMP = ERSUMP+ERR(IK)
    IF (MM(1).NE.2) THEN
      FP = FP+ERR(IK)*ERR(IK)
      SSQ = SSQ+RUNOFF(IK)*RUNOFF(IK)
    ENDIF
78  CONTINUE
C
C  END OF MONTH
C

SUM = SUM + SUMM
SUMP = SUMP + SUMPM
ERRM = SUMPM-SUMM
TERRM = SUMP-SUM
IF (IFIN.NE.0) THEN
  TRAIN = TRAIN + SRRAIN
  SEVAP = SEVAP + SUMEM
  SEPRED = SEPRED + SUMEP
  DO 80 I=1,31
    IF (RAIN(I).LE.0.0) THEN
      DIROP(I)= 0.0
    ELSE
      DIROP(I)= 100.0*DIRO(I)/RAIN(I)
    ENDIF
    IF (SRRAIN.EQ.0.0) THEN
      SDIROP = 0.0
    ELSE
      SDIROP = 100.0*SDIRO/SRRAIN
    ENDIF
80  CONTINUE
ENDIF
IF (MM(1).EQ.2) FP = FP+ERRM*ERRM
IF (IFIN.NE.0) THEN
  ITD = ITD+ID
  TD = FLOAT(ITD)
  IF (SUMM.EQ.0.0) THEN
    ERRMPC = 0.0
  ELSE
    ERRMPC = 100.0*ERRM/SUMM
  ENDIF
  IF (SUM.EQ.0.0) THEN
    TERMPC = 0.0
  ELSE
    TERMPC = 100.0*TERRM/SUM
  ENDIF
  IF (MM(1).EQ.2) THEN
    TD = FLOAT(IM)
    SSQ = SSQ+SUMM*SUMM
  ENDIF
  IF (TD.NE.0) THEN
    FO = SSQ-(SUMM*SUMM)/TD
  ELSE
    FO = SSQ
  ENDIF
  IF (FO.NE.0.0) THEN
    EFFCY = (FO-FP)/FO

```

```

ELSE
  EFFCY = 0.0
ENDIF
IF (MM(4).LE.0) WRITE(8,204)IDATE,(I,RAIN(I),
&EVAP(I),EPRED(I),DEF(I),GSTORE(I),DIROP(I),RUNOFF(I),
&PRUN(I),ERR(I),ERPC(I),I=1,ID)
IF (MM(4).LT.1) WRITE(8,207)SRAIN,SUMEM,
&SUMEP,SUMM,SUMP,ERRM,ERRMPC,IDATE,SUM,SUMP,FF,EFFCY,CS,DC,GS,
&SDIROP,TRAIN,SEVAP,SEPRD
IF (MM(4).EQ.1) THEN
  IF (IM.NE.1) THEN
    IM1 = IM1+1
    IMM = IM1/12
    IF (IM1-(IMM*12).EQ.0) WRITE(8,208)
  ELSE
    WRITE(8,208)
  ENDIF
  WRITE(8,209)IDATE,SRAIN,SUMEM,SUMEP,SUMM,SUMPM,
&ERRM,CS,DC,GS,ERRMPC,TRAIN,SEVAP,SEPRD,SUM,SUMP,TERRM,
&TERMP,FF,EFFCY
ENDIF
F=FP
IF(IFIN.GE.1.OR.MSR(3).EQ.1) THEN

  CALL STAT

ENDIF
ENDIF

IF (IDATE.NE.MM(3)) GOTO 18

C
C
C  END OF MODEL RUN

DIF=SUMP-SUM
FF=FP
F=FP
PRINT*, 'F = ',F

C
IF(IFIN.EQ.0.AND.III.EQ.1.AND.NUM.GT.0)
&WRITE(8,210) SUM,(PNOM(I),I=1,NUM)
IF(NUM.GT.0.AND.IFIN.EQ.0)WRITE(8,212)F,SUMP,DIF,(Y(I),I=1,NUM)
IF(IFIN.GT.0.AND.MM(4).EQ.IDATE) WRITE(8,214)CS,DC,GS
IF(IFIN.GT.1) THEN
DO 76 K=1,NUM
76  WRITE(8,286)K,PNOM(K),YI(K),F
ENDIF
RETURN

C
C  Error RETURN - Incompatible data and MODELLING FREQUENCIES.
C  ABORTS MAIN PROGRAM.
C

286 FORMAT(10X,I2,A8,F12.5,5X,'F = ',F12.6)
200 FORMAT(1X,(1X,F9.5,))
203 FORMAT(1H0,5X,'DATE: ',I2,2X,I4,5X,'RUNOFF ',
&F6.3,5X,'PRUN ',F6.3)
204 FORMAT(1H0,5X,'MONTH: ',I4//1X,'DAY',5X,'RAIN',3X,'EVAP',3X,
&'EPRED',2X,'DEF',3X,'GSTORE',5X,'DIR RO%',4X,'RUNOFF',6X,'PRUN',
&3X,'PRUN-RUNOFF',4X,'ERROR %'//31(2X,I2,2X,

```

```

&5F7.1,F12.2,F11.3,F10.3,F11.3,F14.1/))
207 FORMAT(1H,'TOTALS',F6.1,F7.1,F7.1,27X,2F10.3,F11.3,F14.1//9X,
&'RAIN',3X,'EVAP',3X,'EPRED',30X,'RUNOFF',6X,'PRUN',4X,'PRUN-RUNO',
&'OFF',4X,'ERROR %'//1X,'AT END OF MONTH ',I4,' CUMULATIVEVALUES',
&' ARE: ', ' FLOW =',F8.1,' PREDICTED =',F8.1,' VARIANCE=',F13.2,
&' EFFICIENCY =',F7.3/21X,'STORE CONTENTS ARE: CS =',F6.2,' DC=',
&F6.1,' GS =',F7.1,' - DIR.R.O. =',F5.2,'% OF RAIN'/
&/21X,'CUMULATIVE VALUE OF RAIN =',F10.1,5X,'EVAP =',F10.1,
&5X,'EPRED =',F10.1)
208 FORMAT(1H1/1X,'MONTH RAIN EVAP EPRED RUNOFF PRUN PRU',
&'N-RUNOFF CS DC GS ERROR % VARAINCE ',
&'EFFICIENCY '/')
209 FORMAT(/,1X,I5,5F8.1,2X,F6.1,4X,F5.2,2F6.1,F14.1/7X,6F8.1,29X,
&F6.1,4X,F13.2,4X,F7.3)
210 FORMAT(1H0,30X,'VARIANCE, FLOW AND ACTIVEPARAMETERS'/31X,36(1H*),
&/1X,'SUM=',F10.4//6X,'FF',7X,'SUMP',5X,'SUMP-SUM',10(1X,A8)/,
&31X,10(1X,A8))
212 FORMAT(1H ,3F12.5,4F9.5/31X,5F9.5)
214 FORMAT(1H0//20X,'FINAL STORE CONTENTS',6X,'CS =',F7.4,6X,
&'DC =',F8.4,6X,'GS =',F8.4)
604 FORMAT(5(F13.6,1X)/5(F13.6,1X))
END

```

SUBROUTINE READER

```
INCLUDE "COMMON.DAT"
```

```

IF(IM.EQ.1) ISTART=0
1 READ(4,END=99) IDATE,ID
READ(4)(RAIN(J),RUNOFF(J),EVAP(J),J=1,ID)
IF(IDATE.EQ.MM(2)) ISTART=1
IF(IDATE.EQ.MM(3))REWIND 4
IF(ISTART.EQ.1) RETURN
GOTO 1

99 RETURN
END

```

SUBROUTINE STAT

```
INCLUDE 'COMMON.DAT'
```

```

C *** Subroutine to calculate MAX. and MIN. of predicted and observed flows
IF(MM(2).EQ.IDATE)JDATE=IDATE
IF(MM(3).EQ.IDATE)KDATE=IDATE
IF(IM.GT.1)GOTO 11
OMAX=0.0
PMAX=0.0
OMIN=1000.0
PMIN=1000.0
11 DO 19 J=1,ID
IF(RUNOFF(J).LE.OMAX) GO TO 13
OMAX=RUNOFF(J)
PMAXO=PRUN(J)
KJ=J
IID=IDATE
13 IF(PRUN(J).LE.PMAX) GO TO 15
PMAX=PRUN(J)
OMAXP=RUNOFF(J)
IJ=J

```



```

      IDD=IDATE
15  IF(RUNOFF(J).GE.OMIN) GO TO 17
      OMIN=RUNOFF(J)
      PMINO=PRUN(J)
      JJJ=J
      IIDD=IDATE
17  IF(PRUN(J).GE.PMIN) GO TO 19
      PMIN=PRUN(J)
      OMINP=RUNOFF(J)
      JI=J
      IDI=IDATE
19  CONTINUE
      IF(MM(3).EQ.IDATE)GOTO 20
      RETURN
20  IF(MM(1).NE.2) RM=ITD
      IF(MM(1).EQ.2) RM=IM
      IFFCY=0
      EFFCY=(FO-F)/FO
      IF(EFFCY.LT.0.0) IFFCY=1
      IF(IFFCY.EQ.1) EFFCY=(-1.0)*EFFCY
      CORR=SQRT(EFFCY)*100.0
      IF(IFFCY.EQ.1) THEN
      CORR=(-1.0)*CORR
      EFFCY=(-1)*EFFCY
      ENDIF
      ERTOTL=100.0*(SUMP-SUM)/SUM
      FRM=F/RM
      IF(FRM.LT.0.0) FRM=(-1.0)*FRM
      PE=SQRT(FRM)/(SUM/RM)
      RATIO=SEPPRED/SEVAP
      WRITE(8,115) MM(2),KDATE,TRAIN,SUM,SUMP,OMAX,PMAXO,KJ,IID,
&PMAX, OMAXP,IJ,IDD,OMIN,PMINO,JJJ,IIDD,PMIN,OMINP,JI,
&IDI,ERTOTL,FO,F,EFFCY,PE,CORR,SEVAP,SEPPRED,RATIO
23  RETURN
C
115  FORMAT(///3X,'MONTH ',I4,' TO MONTH ',I4,' INC.'//10X,'TOTAL RAIN'
& ', 'FALL (MM)',9X,F12.3//10X,'TOTAL FLOW (MM)',5X,'OBSERVED',F12.5
& '/29X,'PREDICTED',F12.5//12X,'MAX FLOW (MM)',5X,'OBSERVED',F12.5,
& '/29X,'PREDICTED',F12.5//33X,'DATE - ',I2,I4/29X,'PREDICTED',
& F12.5/29X,'OBSERVED',F12.5//33X,'DATE - ',I2,I4//12X,
& 'MIN FLOW (MM)',5X,'OBSERVED',F12.5//29X,'PREDICTED',F12.5/,
& 33X,'DATE - ',I2,I4/29X,'PREDICTED',F12.5//29X,'OBSERVED',F12.5,
& /33X,'DATE - ',I2,I4/
& /12X,'ERROR IN TOTAL DISCHARGE =',F12.3,'% '//1X,'INITIAL VARIANCE
& =',F12.3/1X,'FINAL VARIANCE =',F12.3/1X,'EFFICIENCY-EXPLAINED',
& ' VARIANCE =',F12.4//1X,'COEFFICIENT OF VARIATION =',F12.3//1X,
& 'CORRELATION COEFFICIENT =',F12.3,'% ',///20X,'TOTAL EO =',F10.3,
& 5X,'TOTAL EPRED =',F10.3,5X,'RATIO EPRED/EO =',F10.3)
120  FORMAT( ' STOP...EFFICIENCY IS LESS THAN 0.0')
      END

```

Control program input data:-

INPUT.DAT

```

0 1 0 1 0 0 0 0      ****Call Rosenbrock and Stat
THESIS MODEL, 9 PAR.  ****Model Title printed
**** Balquhiddar 28 **** ****Catchment Title printed
***** 0584 TO 1286 ***** ****Data period printed
0 0584 1286 2      ****Data frequency (0=daily, 2=monthly), data period, monthly
                                printout
4 09 20 1.5000      ****Active parameters, total parameters, iterations, step
SS      1  2.06420  0.0001  10.0 ****Active parameter, parameter & bounds
RC      2  0.80000  0.10    0.98
RDEL    3  16.54177  1.0    28.0
RX      4  2.48837  1.0    5.0
RK      0  0.22385  0.01   1.0
FC      0  0.80469  0.40   1.0
GSU     0  82.61760  70.0   150.0
GSP     0  1.08474  1.0    5.0
GDEL    0  17.78181  0.1    25.0

```

COMMON.DAT

```

CHARACTER*30 JJ(3)
CHARACTER*6 PNames(30),PNOM(30)

COMMON/B1/IM,ID,III,IDATE,MM(6),RUNOFF(31),RAIN(31),ICAT
COMMON/B2/F,N,NN,IFIN,IGR,kk(30),PNOM,Y(30),YI(30),YK(30)
COMMON/B3/EVAP(31),IFDC,SCALES
COMMON/B4/L,X(30),Z(3,30),A(30,30),PNames
COMMON/B5/SUMM,SUMPM,SRAIN,NR,NP,MC,MD,MA(9),MB(9),PRUN(31)
COMMON/B6/IDATAR(685),ISUMER(4170),ITD,ISTD
COMMON/B7/ISCAT(51,51),SCAT,SCT,IHPF,CON
COMMON/B9/SSQ,SUM,SUMP,TRAIN,SEVAP,SEPRD,TD,FO
COMMON/B10/MSR(8)
COMMON/A1/ISD(9),IFD(9),I04
COMMON/M7/P(30,31),PBAR(30),PSTAR(30),PDSTAR(30),YSI(31),PI(30),
&SPY(30),KC(30),YY(30),B(30),C(30)
COMMON/extra/ims,imw,consafe

```

N.B. This is NOT the FORTRAN77 code for the package version of HYRROM which has different and restrictive settings for parameters, limits are ± 0.5 of parameter initial value, and parameters cannot be set outside bounds specified in the manual.

Appendix 3.

```

SUBROUTINE MUTATE (NL,NM,BKORRL,DELTAS,
& DELTAI,DELTAP,N,NS,NP,X,S,P,D)

```

```

LOGICAL BKORRL

```

```

DIMENSION X(N),S(NS),P(NP)

```

```

EXTERNAL SEED,RANDOM

```

C

C Handles the random alteration of the strategy variables
C and the object variables.

C

```

DS = GAUSSN(DELTAS,D)
DO 1 I=1,NS
IF(I.EQ.1) CALL SEED(INT(D))
IF(I.GT.1) THEN
CALL TIMEDEL(I,Q)
I4A=I4
CALL GETTIM(IH1,IH2,IH3,I4)
IF(I4.EQ.I4A) I4=I4+5
CALL SEED(I4)
ENDIF
CALL RANDOM(R1)
R1=R1*100.0
1 S(I) = S(I)*EXP(DS+GAUSSN(DELTAI,R1))
DO 2 I=1,N
IF(I.EQ.1) CALL SEED(INT(R1))
CALL TIMEDEL(I,Q)
I4A=I4
CALL GETTIM(I1,I2,I3,I4)
IF(I4A.EQ.I4) I4=I4+5
CALL SEED(I4)
CALL RANDOM(R2)
R2=R2*100.0
2 X(I)=GAUSSN(S(I),R2)
IF (.NOT.BKORRL). RETURN
DO 3 I=1,NP
if(i.eq.1) call SEED(INT(R2))
CALL TIMEDEL(I,Q)
I4A=I4

```

```
      CALL GETTIM(I1,I2,I3,I4)
      IF(I4.EQ.I4A) I4=I4+5
      CALL SEED(I4)
      CALL RANDOM(R3)
      R3=R3*100.0
3     P(I) = P(I)+GAUSSN(DELTAP,R3)
      CALL DREHNG(NL,NM,N,NP,X,P)
      D=R3
      RETURN
      END
```

```
      SUBROUTINE TIMEDEL(I,Q)
      Q=0.0
      DO 10 J=1,100
      DEL = I*EXP(2.0)
      Q = Q + DEL
10     Q= 1.0*EXP(2.0)
      RETURN
      END
```